Journal of the
OPTICAL SOCIETY
of AMERICA

VOLUME 31 APRIL, 1941 NUMBER 4

Application of the Electron Microscope to the Study of Photographic Phenomena*

C. E. HALL AND A. L. SCHOEN
Eastman Kodak Company, Rochester, New York

(Received February 1, 1941)

An electron microscope having a resolving power of less than 50A has been built and used to examine photographic emulsions. When photographic crystals are exposed to an electron beam, silver ions move within the crystals and metallic silver is produced. Also the crystals disintegrate, cracks and holes developing. Pictures have been made showing very small particles of silver formed within silver bromide crystals when they are exposed to an intense light source. Silver reduced from silver bromide by the common chemical developers has been found to have a filamentary structure rather than the "cokelike" form as previously believed. Physical development produces clumps of closely packed silver particles. The smallest photographic crystals (Lippmann crystals), when developed in the common developers, are converted into single threads or ribbons of silver.

I. APPARATUS AND TECHNIQUE

A MAGNETIC electron microscope of the type developed by Burton[1] and co-workers† at the University of Toronto has been constructed. The basic structure is similar to that described by Prebus and Hillier[2] to which has been added a locking chamber for inserting the object into the vacuum and a magazine camera holding twelve photographic plates. A photograph of the instrument is shown in Fig. 1(a) and a diagram of the image formation in Fig. 1(b). Electrons from the tip of a hot tungsten filament are accelerated through a high potential and pass through a hole in the anode. By means of the condenser lens, the electron beam can be collimated or focused on the object. Electrons passing through or around the object are focused by the objective lens to form a magnified image of the object on the first fluorescent screen. A small hole in this screen allows a selected part of the first image to be further magnified by the projector lens. The final image is focused visually on a fluorescent screen which can be moved out of the way to allow the electrons to strike the photographic plate. Exposure times are from 1 to 10 seconds. The interior of the instrument is evacuated to less than 10^{-4} mm of mercury.

The accelerating voltage is obtained from a conventional 70-kv rectifier circuit operating from the 110-volt, 60-cycle supply. This voltage is maintained sufficiently constant by the use of commercial type constant voltage transformers in the primary of the high voltage transformer and large filter condensers (0.5μf) in the secondary. The magnetic lenses are operated from the 110-volt d.c. supply through an electronic voltage regulator.

* Communication No. 790 from the Kodak Research Laboratories.

[1] E. F. Burton, J. Hillier, and A. Prebus, Phys. Rev. **56**, 1171 (1939).

† Under the direction of Professor Burton, construction of electron microscopes was begun in 1935 by W. H. Kohl and C. E. Hall, and continued by A. Prebus and J. Hillier.

[2] A. Prebus and J. Hillier, Can. J. Research **A17**, 49 (1939).

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

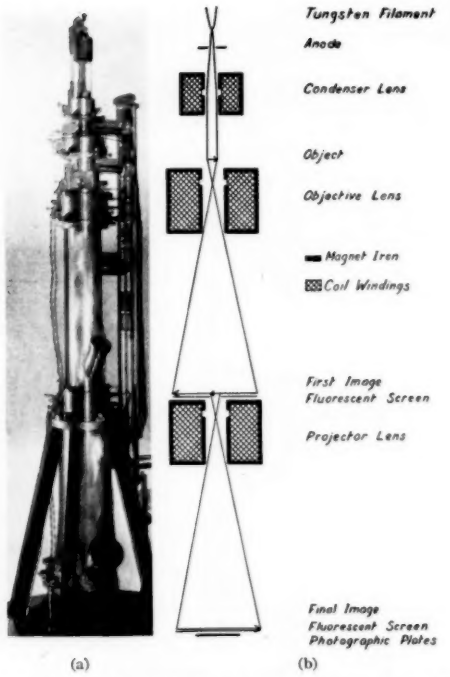


FIG. 1. (a) Photograph of electron microscope. (b) Diagram of image formation.

The electron microscope is usually operated at magnifications from 15,000 to 25,000 diameters. This is obtained by magnifying 100 times in the first stage and about 200 times in the second. Most pictures thus obtained contain enough sharp detail to permit further useful photographic enlargements resulting in magnifications from 50,000 to 100,000 diameters. The resolving power is estimated at less than 50A. In common with other electron microscopes of this type, the angle of illumination at the object is reduced to about 3×10^{-3} radian in order to minimize the effects of spherical aberration in the magnetic lenses.[3] This results in a large depth of focus. Von Ardenne[4] has calculated that, under practical working conditions, the depth of focus is about 2 microns.

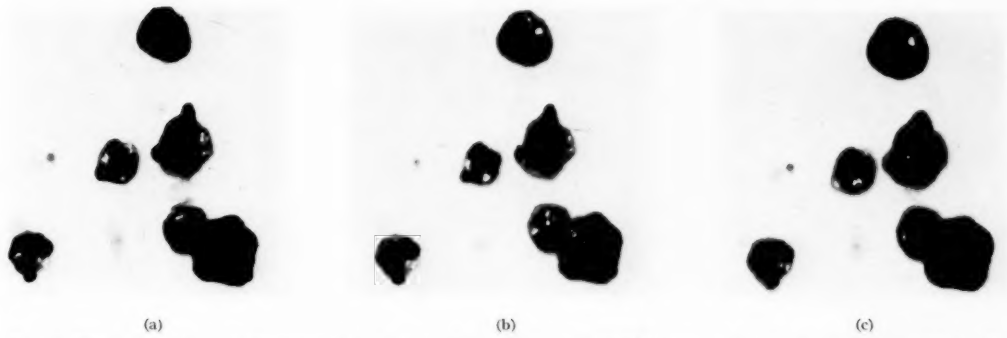
High resolving power can be attained only with very thin (100–500A) objects.[5] Object supports are made from a 1 to 2 percent solution of cellulose nitrate in amyl acetate.[6] When a drop of this solution is allowed to spread over the surface of a tray of water, the amyl acetate evaporates leaving behind a thin film of cellulose nitrate, estimated at about 100A thick. This film is then picked up over a fine wire mesh, or a thin platinum disk with a hole punched in the center. A drop of dilute photographic emulsion, when placed on this support and allowed to evaporate, forms an object suitable for examination in the electron microscope.

II. Effect of Electron Exposure on Photographic Crystals

Silver halide crystals used in photographic emulsions possess the property of becoming ionic conductors when silver ions within the crystal lattice are released from their fixed positions by the absorption of energy. The ability of these silver ions to move about within the crystal lattice, which is of fundamental importance to the formation of the photographic latent image,[7] results in an unusual behavior of photographic crystals when they are subjected to the intense electron beam necessary for their observation in the electron microscope. The crystals are, at first, opaque to the electrons but, after continued exposure, they begin to show transparent holes and cracks due to migration

FIG. 2. Silver bromide crystal exposed to intense electron beam. ×36,000.

[3] B. von Borries and E. Ruska, Zeits. f. tech. Physik **20**, 225 (1939).
[4] M. von Ardenne, Naturwiss. **28**, 248 (1940).
[5] M. von Ardenne, Zeits. f. Physik **111**, 152 (1938).
[6] H. Ruska, Naturwiss. **27**, 287 (1939).
[7] J. H. Webb, J. App. Physics **11**, 18 (1940).



(a) (b) (c)

FIG. 3. Silver chloride crystals exposed to intense electron beam. ×34,000, approx. (a) 2-min. exposure. (b) 4-min. exposure. (c) 8-min. exposure.

of the silver ions. Such a crystal of silver bromide is shown in Fig. 2. The appearance of the crystals gradually changes while the electron exposure continues until a stable state is reached. This effect has been described by von Ardenne.[8] A series of pictures in Figs. 3(a), 3(b), and 3(c) shows consecutive steps in this print-out process for a group of silver chloride crystals. Some places in a crystal which are at one time transparent, become at a later time opaque, indicating that the material of the crystal is rendered quite mobile by the action of the electron beam. In Figs. 3(b) and 3(c), a semitransparent substance can be seen creeping out from the crystal boundaries. Metallic silver and other stable materials do not show these phenomena.

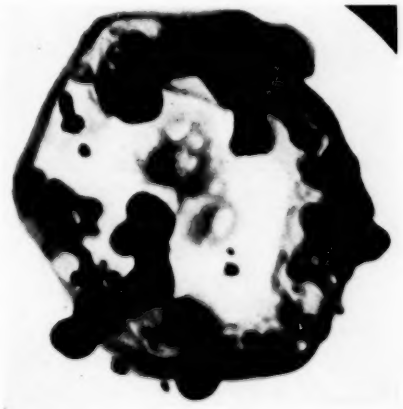


FIG. 4. Silver bromide crystal exposed to electron beam. Fixed. ×40,000.

[8] M. von Ardenne, Zeits. f. angew. Phot. **2**, 14 (1940).

III. Reduction of Silver by Electron and Light Exposure

It has been established that certain points of imperfection in the silver halide crystal lattice, associated with the presence of sulfur,[9] are able to trap and collect migratory silver ions. When enough silver has been concentrated at these points, the silver halide crystal is rendered susceptible to the reducing action of photographic developers. Trillat and co-workers[10,11] have demonstrated the presence of silver within the exposed silver bromide crystals by x-ray and electron diffraction studies.

Electron micrographs have been made showing silver reduced from silver bromide crystals by exposure to electrons. The crystals were exposed to the electron beam while they were observed in the electron microscope. They were then taken from the electron microscope and bathed in a sodium thiosulfate solution. This dissolved and removed the silver bromide but left unaffected any silver reduced by the electron exposure. A crystal treated in this manner is shown in Fig. 4. It is to be noted that the reduced silver does not cover the area of the crystal uniformly but tends to gather about certain preferred areas. The outline of the crystal corresponding to the halos in Fig. 3(c) still remains, indicating that these halos are silver, at least in part. The reduction of silver by the electron beam is accompanied by the release of halogen, which

[9] S. E. Sheppard, Phot. J. **65**, 380 (1925).

[10] J. J. Trillat and H. Motz, J. de phys. et rad. **7**, 89 (1936).

[11] J. J. Trillat and R. Merigoux, J. de phys. et rad. **7**, 497 (1936).

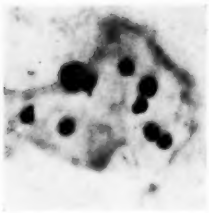


FIG. 5. Photolytic silver particles. ×25,000.

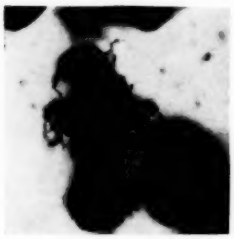


FIG. 6. Silver bromide crystals. Partially developed. Not fixed. ×20,000.

FIG. 7. Silver produced by chemical development (hydroquinone). ×25,000.

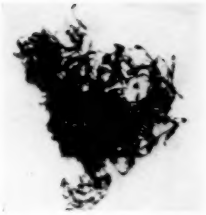


FIG. 8. A silver grain developed in amidol. ×20,000.

would readily escape to the evacuated interior of the electron microscope.

Latent image particles produced by exposures to light of the order used in practical photography have not been definitely identified from the many other small particles present in a photographic emulsion. When very high exposures are used, enough silver atoms accumulate at the latent-image centers so that they can be readily seen and distinguished from the background. Such particles of silver, with a shadowy outline of the original crystal in the background, are shown in Fig. 5. Pictures of this kind are made by exposing silver bromide crystals to an intense light source, after which the silver bromide is removed with sodium thiosulfate, leaving behind only the concentrated silver centers. These centers cannot be found if the crystals are unexposed and they increase in size as the exposure is increased. In many cases, they have regular shapes indicating that they are crystalline silver.

IV. Appearance of Developed Silver

In Fig. 6 is shown a group of silver bromide crystals which have been immersed in a common developer solution for a short time. Some of the silver bromide crystals have not been affected appreciably by the developer. The filamentary structure is silver reduced from silver bromide by the developer and is similar to the type of structure previously noted by von Ardenne.[8] The filamentary character of developed silver is shown more clearly in Figs. 7 and 8, where silver bromide crystals have been developed and then fixed to remove any unreduced silver bromide. Studies with optical microscopes have resulted in a generally accepted view that developed silver was a "cokelike" substance. It is now apparent that it is more like a mass of tangled threads and ribbons.

The grains in Figs. 7 and 8 were developed on the electron microscope object holder. Although they show the kind of silver deposit of which a developed silver grain is composed, they are not the same in general appearance as the grains developed on a photographic plate. In grains fully developed on a photographic plate, the filaments are so closely packed that most of the grain is completely opaque with filamentary structure in evidence only around the edges. Occasional filaments reach out far beyond the original crystal boundaries so that they may penetrate a neighboring crystal and render it developable.

With the exception of pure paraphenylenediamine, all of the common chemical developers produce this typical filamentary structure with some secondary differences. Metol and amidol, for example, produce very fine filaments, while hydroquinone produces coarse filaments. The extent to which the filaments can wander away from the original crystal is influenced by the rigidity of the surrounding gelatin.

Figure 9 shows what was originally a single silver bromide crystal which has been developed for a very short time and then fixed. A faint outline can be seen in the background where the original crystal has left its trace in the gelatin. There are many small particles of silver present, confirming the view that development starts at certain preferred centers, presumably at the latent-image centers.

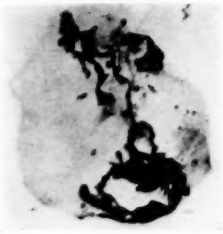


FIG. 9. Partially developed silver bromide crystal. Fixed. ×20,000.

FIG. 10. Silver produced by physical development. ×20,000.

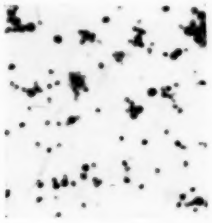


FIG. 11. Undeveloped Lippmann crystals. ×25,000.

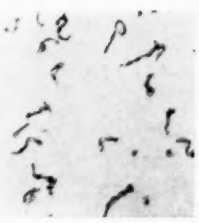


FIG. 12. Developed Lippmann crystals. ×25,000.

A photographic image can be produced from an exposed emulsion by another method known as physical development. By this method, the silver bromide is first dissolved and silver is then deposited from solution on the photolytic particles. This type of development produces an image of very fine grain. The kind of silver deposit produced by a physical development method devised by Odell[12] is shown in Fig. 10. These clumps of silver are in marked contrast to the filaments of chemically developed silver. Paraphenylenediamine developer produces a deposit similar to Fig. 10.

Lippmann emulsions (Fig. 11), containing the smallest photographic crystals, show some particularly interesting effects. When developed in common chemical developers, each small crystal grows into a single filament which is much longer and thinner than the original crystal. The resulting wormlike structure of a developed Lippmann emulsion is shown in Fig. 12.

SUMMARY

An electron microscope having a resolving power of less than 50A has been built and used to examine photographic emulsions. When photographic crystals are exposed to an electron beam, silver ions move within the crystals and metallic silver is reduced from the silver halide. The crystals disintegrate, showing cracks and holes. Pictures have been made showing very small particles of silver formed within silver bromide crystals when they are exposed to an intense light source. Silver reduced from silver bromide by the common chemical developers has been found to have a filamentary structure rather than the "cokelike" form as previously believed. Physical development produces clumps of closely packed silver particles. The smallest photographic crystals (Lippmann crystals), when developed in the common developers, are converted into single threads or ribbons of silver.

ACKNOWLEDGMENTS

We are indebted to Professor E. F. Burton and co-workers of the University of Toronto for making available the constructional details of their electron microscope, and to Dr. C. J. Staud, Dr. E. E. Jelley, and many other members of these Laboratories for contributing valuable suggestions and discussion.

[12] A. F. Odell, Brit. J. Phot. **80**, 286–87, 303 (1933).

Electro-Optical Field Mapping

HANS MUELLER
George Eastman Research Laboratories of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received February 14, 1941)

As a result of the discovery of an unusually large electro-optical effect in certain colloidal solutions of bentonite it has become possible to study and measure inhomogeneous electric fields by an optical method which is analogous to that of photoelasticity. Observations with crossed Polaroids show two orthogonal systems of lines. The isoclinic lines determine the field direction, the isochromatic lines give the field intensity in any point, of a plane field. The method employs a.c. potentials of less than 200 volts.

IT is well known that most liquids become doubly refracting under the influence of an electric field E. The electro-optical birefringence is uniaxial with the optical axis in the field direction. The ordinary and extraordinary components of a light wave traveling normal to the field suffer a relative phase shift

$$\varphi = 2\pi l B E^2, \qquad (1)$$

where l is the length of the optical path in the liquid and B is the Kerr constant. From measurements of φ and of the direction of the optical axis it is, therefore, possible to infer the magnitude and direction of the field intensity in any point of a plane field distribution. In principle this electro-optical method of field mapping is simpler than the analogous problem of elastic strain analysis by photoelasticity. In the latter the optical effects are related to a tensor quantity, in the former they serve to find a vector. In practice, however, the electro-optical investigations encounter greater difficulties. This is due to the fact that, while large phase shifts amounting to several multiples of 2π can readily be produced by means of elastic stresses, it has until recently been impossible to create electro-optical double refraction of a similar magnitude. Even in nitrobenzene, though its Kerr constant ($B = 10^{-5}$ e.s.u.) is a thousand times larger than that of most other liquids, the production of a phase shift of 2π requires a field of an intensity approaching the electrical breakdown strength and an excessively long light path in the liquid. Since field mapping becomes very laborious when the phase shifts are small, the electro-optical method can be made practical only if substances with larger Kerr constants can be employed.

From the work of Procopju,[1] Björnståhl,[2] and Errera, Overbeek and Sack[3] it is known that many suspensions, colloids and proteins[4] possess Kerr constants which exceed that of nitrobenzene by factors ranging from 10^3 to 10^7. Unfortunately many of these substances are not suitable for our purpose for one or both of the following reasons: They are not sufficiently transparent and the length of the active light path is limited. The Kerr effect in colloids approaches a saturation value and although the Kerr constant is large for small fields the maximum birefringence frequently is smaller than that attainable in nitrobenzene. Due to their electrical conductivity the colloids are heated in moderately strong fields. The heating creates convection currents which give rise to streaming birefringence and disturb the image formation.

In the course of a recent investigation[5] of the electro-optical properties of colloids, carried out in collaboration with Dr. B. W. Sakmann, we have discovered certain solutions of bentonite in which the above limiting factors are minimized to such an extent that the electro-optical method of field mapping becomes practical. These aqueous solutions are obtained by dilution of gels of yellow bentonite. The gels were prepared by Professor E. A. Hauser[6] and Dr. D. S. le Beau in the laboratories of chemical engineering of

[1] S. Procopju, Ann. de physique **10**, 213 (1924).
[2] Y. Björnståhl, Thesis, Upsala (1924).
[3] J. Errera, W. Overbeek and H. Sack, J. Chim. Phys. **32**, 681 (1935).
[4] M. A. Lauffer, J. Am. Chem. Soc. **74**, 147 (1939).
[5] H. Mueller, Phys. Rev. **55**, 508, 792 (1939); F. J. Norton, Phys. Rev. **55**, 668, 1939; H. Mueller and B. W. Sakmann, Phys. Rev. **56**, 615 (1939).
[6] E. A. Hauser and D. S. le Beau, J. Phys. Chem. **42**, 1031 (1938); **43**, 1037 (1939); **45**, 54 (1941).

Massachusetts Institute of Technology. Only such sols are suitable in which the average diameter of the clay particles is less than 20 millimicrons. The gels are gradually diluted until the spontaneous birefringence disappears and the relaxation time of the streaming birefringence is reduced to about 1/100 sec. This occurs at a concentration of approximately 2 percent by weight. The resulting yellowish liquid is transparent in thicknesses up to 10 cm. The Kerr constant for 60-cycle a.c. fields is of the order of magnitude $B=10$ e.s.u. The saturation value of the birefringence is so large that for a light path of 1 cm the maximum phase shift exceeds the value 2π. The saturation occurs for fields of about 150 volts/cm. For fields of 100 volts/cm the heating effects become noticeable after a period of about one minute. They are not disturbing since the electro-optical effects can be photographed with an exposure time of one second. The values of the Kerr constant and of the saturation birefringence vary with the particle size, the sol concentration, the temperature and the frequency of the electric field. After some experience one is able to adjust the Kerr constant to any convenient value by diluting or concentrating the sol, but the large values of φ can be obtained only with the finely dispersed colloids.

In order to demonstrate the practicability and reliability of the electro-optical method of field mapping by means of these colloids we have studied the well-known field distribution around two parallel cylindrical conductors. If the conductors are oppositely charged the equipotential lines and the lines of force are circles, as shown in Fig. 1. These lines are not recorded by the Kerr effect. The electro-optical method furnishes the lines of constant field intensity and the lines of constant field direction. This follows from the following elementary fact: If the Kerr cell is placed between crossed Nicols and the field direction forms an angle Θ with the plane of vibration of the polarizer, the intensity of the transmitted light is proportional to

$$I=\sin^2 2\Theta \sin^2 \tfrac{1}{2}\varphi. \qquad (2)$$

Hence the intensity is zero when $\Theta=0$ or $\Theta=90°$. With an inhomogeneous plane field distribution in the Kerr cell we observe, therefore, a set of dark lines which correspond to the locus of points where the field direction is in the plane of vibration of either the analyzer or the polarizer. The position of these "isoclinic" lines is independent of the wave-length of the light; they always are black. They are not influenced by a change of the voltage between the conductors and are independent of the length of the Kerr cell.

By rotation of the crossed Nicols the isoclinic lines can be shifted so as to pass through any

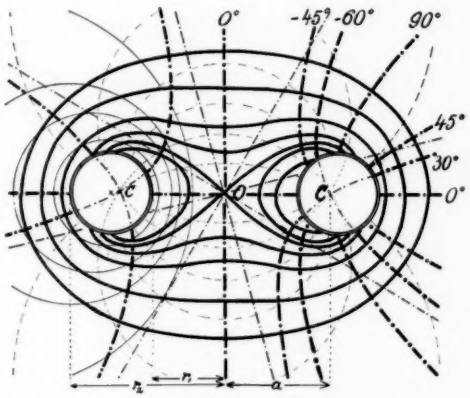


FIG. 1. The electric field around two oppositely charged parallel rods. The thin circles are the equipotential lines and the lines of force (dotted). The thick lines are the lines of constant field intensity and correspond to the isochromatic lines. The hyperbolas (dot-dashed) are the isoclinic lines for various field directions.

arbitrarily chosen point in the field. The fact that an isoclinic line passes through a point P when the plane of vibration of the polarizer is at an angle ψ with the horizontal direction implies that the field in this point forms one of the four angles ψ, $\psi\pm90°$, $\psi+180°$ with the horizontal. It usually is easy to decide which one of the four possibilities is the correct one.

For the field around the parallel rods the isoclinic lines can be calculated and are shown in Fig. 1. They are equilateral hyperbolas. For every Nicol position ψ we obtain two hyperbolas which intersect in the centers C. The asymptotes of one hyperbola form with the horizontal the angles $\frac{1}{2}\psi$ and $\frac{1}{2}(\psi-\pi)$. The asymptotes of the second hyperbola bisect the angles between those of the first hyperbola. For the case $\psi=0°$ or $\psi=90°$ one hyperbola degenerates into the two axes of symmetry.

According to Eq. (2) the intensity of the transmitted light vanishes also when the phase shift φ is a whole multiple of 2π. For a fixed length of the optical path this occurs when the field intensity assumes one of a series of critical values $E_1, E_2, \cdots, E_i, \cdots$. For our colloids Eq. (1) is not valid but is to be replaced by a more complicated relation of the general form

$$\varphi = 2\pi l f(E^2), \qquad (1')$$

where the function f depends on the characteristics of the colloid. Hence the critical field intensities have to be determined empirically for every solution. For an inhomogeneous field the observation through crossed Nicols will therefore show a second set of dark lines which are the loci of points where the field intensity has one of the critical values. These lines are called "isodynamical" or "isochromatic." The first adjective is borrowed from the nomenclature used in terrestrial magnetism. The second is used in photoelasticity and is justified by the fact that in observation with white light these lines are colored because the phase shift depends on the wave-length of the light.

The isochromatic lines are shifted and can be made to pass through any arbitrary point by changing the potentials of the electrodes. To determine the field intensity in a point one makes use of the law which states that the multiplication of the potentials of all electrodes by a common factor alters the field intensity in every point by the same factor. By following the changes of the isochromatic lines with potential variations and from the color sequence it is easy to assign the proper E_i values to the various isochromatic lines.

For the field around two parallel cylindrical conductors the isochromatic lines are given by the equation

$$r^2 = \cos 2\vartheta \pm (E_0^2/E_i^2 - \sin^2 2\vartheta)^{\frac{1}{2}},$$

where r and ϑ are polar coordinates about the origin 0. E_0 is the field intensity in 0 and has the value $E_0 = V/a \ln [(r_1+a)/(a-r_1)]$, where $a = (r_1 r_2)^{\frac{1}{2}}$ is the distance OC (see Fig. 1) and V is the potential between the rods. For $E_i > E_0$ the isochromatic line consists of two rings near the electrodes; for $E_i = E_0$ these rings merge to form a figure 8 with a normal intersect at 0. For $E_i < E_0$ the isochromatic lines form a single ring which approaches circular shape for small values of E_i.

The isochromatic lines are independent of the position of the crossed Nicols. If the latter were rotated at a high speed the isoclinic lines could not be observed and only the isochromatic lines would appear. The same result is accomplished by using circular polarizers, i.e., by sending the light successively through a polarizer, $\frac{1}{4}\lambda$ plate, Kerr cell, $-\frac{1}{4}\lambda$ plate and an analyzer. For this arrangement the transmitted intensity is proportional to $\sin^2 \frac{1}{2}\varphi$ and only the isochromatic lines appear.

The isochromatic lines are orthogonal trajectories of the isoclinic lines. This is true if the electric field has no space charges. To prove this law one takes into account that the isochromatic lines are given by the relation

$$c(x, y) = (\partial\Phi/\partial x)^2 + (\partial\Phi/\partial y)^2 = \text{const.},$$

where Φ is the potential of the field. The isoclinic lines satisfy the equation

$$i(x, y) = \frac{\partial\Phi/\partial y}{\partial\Phi/\partial x} = \text{const.}$$

From Laplace's equation $\Delta\Phi = 0$ it follows

$$\frac{\partial c}{\partial x}\frac{\partial i}{\partial x} + \frac{\partial c}{\partial y}\frac{\partial i}{\partial y} = 0$$

and this is the orthogonality condition for c and i. The relation between the isoclinic and isochromatic lines is therefore analogous to the relation between the lines of force and the equipotential lines. Although the electro-optical observations do not furnish the latter for the field under investigation, the observed lines may represent the lines of force and the equipotential lines of an entirely different electric field problem. In the example of the oppositely charged parallel rods, for instance, the observed lines are identical with the lines of force and the equipotential lines of the field around two rods which carry charges of the same sign. In other cases, where the isoclinic lines begin and end within the field and not at a boundary, the isochromatic lines can be interpreted as the equipotential lines of a field with space charges.

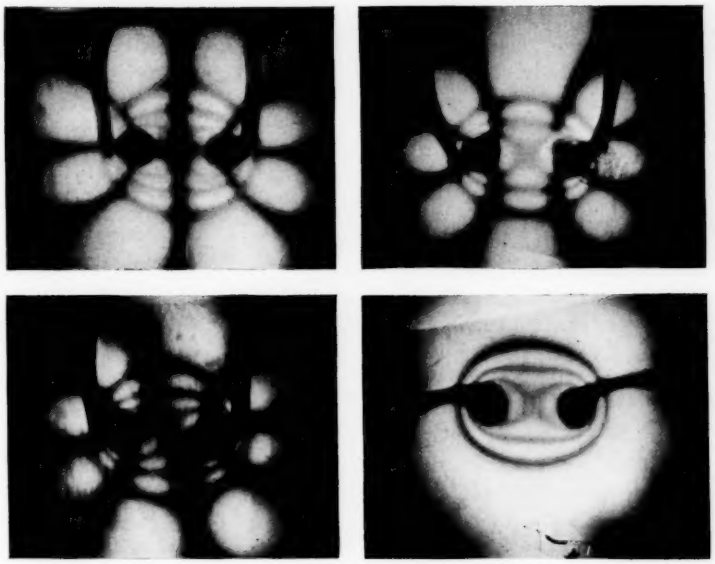


FIG. 2. Electro-optical field maps for the electric field around two parallel rods. The first three maps are made with crossed Polaroids oriented at 0°, 45° and 22.5° to the horizontal, respectively. The last figure gives the isochromatic lines as observed with circular polarizers. The a.c. potential across the rods is 90 volts. (The broad dark lines are shadows of the leads to the electrodes.)

While the success of the experimental method depends primarily on the choice of the colloid, a series of other factors also are important. The colloids will slowly coagulate owing to electrochemical reactions when they are in contact with copper, brass and especially with solder. Electrodes of nickel, nichrome or monel give no difficulties, provided they are clean and are not left in the solution for several hours. Platinum and gold-plated electrodes do not affect the colloids for periods of several weeks. Containers of glass, Bakelite, lucite and several other plastics, held together by Duco cement, serve as Kerr cells. Since the bentonite solutions are electrolytic conductors, with a specific resistance varying between 1000 to 5000 ohm-cm, we are dealing with electric field distributions in a conductor. These fields satisfy Laplace's equation but the boundary conditions differ from the electrostatic case. The electrodes remain equipotential surfaces, but at the surfaces of the colloid the lines of force must be parallel to the surface. Near the top, bottom and side surfaces of the cell the observed lines will, therefore, differ somewhat from the theoretical curves, but if the electrodes are in contact with the windows of the Kerr cell the field distortions at the ends of the electrodes are negligible and the theory for the infinitely long rods is applicable. The leads to the electrodes do not distort the field if the wires are insulated.

The micelles of bentonite are negatively charged. In a d.c. field they migrate to the cathode where they coagulate. This can be prevented only by using alternating potentials between the electrodes, because it has been found impossible to obtain isoelectric solutions with uncharged particles. In ordinary liquids an alternating field $E_0 \cos \omega t$ creates an alternating phase shift $\varphi = 2\pi l B E_0^2 \cos^2 \omega t = \varphi_0(1+\cos 2\omega t)$ and the isochromatic lines become blurred and can be observed only with stroboscopic illumination. In our colloids, however, the particles and the viscosity of the solutions are so large that above a certain frequency the micelles are too sluggish to follow the rapid variations of the field. Nevertheless, they are oriented by the field because the orienting torque is proportional to E^2 and the mean square value of the field intensity does not vanish. For frequencies above 1 kilocycle/sec. these colloids show therefore a Kerr effect, but the electro-optical birefringence is steady. In his thesis Dr. Sakmann[7] has shown that at lower frequencies the phase shift is given by an equation of the form

$$\varphi = 2\pi l\{A(E_0^2) + B(E_0^2) \cos(2\omega t - \alpha)\},$$

[7] B. W. Sakmann, Ph.D. Thesis, Massachusetts Institutue of Technology, 1941.

where the value of B vanishes for high frequencies. For 60-cycle a.c. the ratio B/A between the amplitude of the alternating and the steady part of the phase shift depends on the concentration and temperature of the solution and on the size of the particles. For those sols for which the saturation value of $A(E_0^2)$ is large enough, and which are sufficiently transparent for the purpose of field mapping, this ratio is between 1/10 and 1/100 at room temperature. The isochromatic lines are, therefore, reasonably sharp even for continuous illumination. They are somewhat sharper if we use a light source operated on 60-cycle a.c.

It is advisable to cool the solutions before and between the measurements and to apply the electric fields for short periods only. If the sols get heated above 40°C the convection currents become increasingly annoying, and, since the relaxation time of the particles is reduced in the less viscous medium, the isochromatic lines become diffuse.

The experimental equipment and procedure for electro-optical field mapping are very simple. In our experiments the Kerr cell is a glass trough with parallel windows 6×6 cm wide. The electrodes are rods of nickel, $\frac{3}{16}$ inch in diameter and 5 cm long. They are held in parallel position, $\frac{1}{2}$ inch apart, by two plates of lucite. The a.c. potential is applied over two rubber insulated wires and is adjusted by means of a Variac transformer. Polarizer and analyzer are two sheets of Polaroid. The light source is an a.c.-operated Hg arc. For observations with monochromatic light we use a green filter. The light is made parallel by a first lens, and a second lens focuses the transmitted light into the camera objective. We use Agfa Superpan films with an exposure time of 1 sec. The photographs in Fig. 2 demonstrate the practicability of the method. In a qualitative way the agreement between the observed and the calculated isoclinic and isochromatic lines is satisfactory. Slight deviations are noticeable for the outermost isochromatic line and for the isoclinic lines near the edges of the photographs. These deviations are due to the finite size of the Kerr cell.

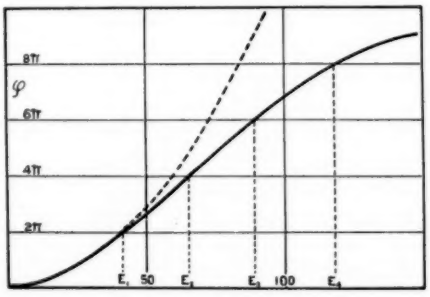


FIG. 3. Typical calibration curve for the electro-optical effect in bentonite solutions for 60-cycle a.c. fields (in volts per cm and for a light path of 5 cm).

A preliminary quantitative test was carried out as follows. Two plane and parallel nickel electrodes, of the same length as the rods and with a plate separation of 1 cm, were inserted in the colloid and the critical values E_i were found from the voltages required to produce extinction between circular polarizers. For the particular colloid used the following values were found

$$E_1=42, \quad E_2=66, \quad E_3=89,$$
$$E_4=118 \text{ r.m.s. volt/cm.}$$

The diagram in Fig. 3 illustrates the deviation of the Kerr effect in colloids from the classical law (1). It should be noted that this type of field dependence is advantageous for our purpose; the isochromatic lines are not crowded together in the region of high field intensities, as they would be if the quadratic law were valid.

This same colloid was used to measure the field intensity in the midpoint between the cylinders. This is done by recording the voltages between the two rods for which the first, second, third and fourth isochromatic lines assume the shape of the numeral 8. These voltages are

$$V_1=41, \quad V_2=64, \quad V_3=87,$$
$$V_4=115 \text{ volt (effective).}$$

The close agreement between the V_i and E_i values, which at first appeared to us quite mysterious, is due to the following coincidence. In our experimental arrangement the distances r_1 and r_2 (see Fig. 1) are 0.40 and 0.87 cm hence $a=(r_1r_2)^{\frac{1}{2}}=0.59$ cm and therefore

$$E_0=V/2.30a \log (r_1+a)/(a-r_1)=1.02\ V.$$

This means that the voltages V_i which are required to produce in the midpoint a field E_0

which coincides with one of the critical values E_i is for our arrangement 2 percent smaller than the numerical values of E_i. The data, therefore, show that the method is quantitative and furnishes the values of the field intensities with an accuracy of about 1 percent.

The calibration of the colloid can also be performed by the photographic method. This is done by studying the field in a cylindrical condenser. From the diameter of the rings shown in Fig. 4 the values of E_i can readily be calculated.

The photographs in Fig. 4 suggest a variety of problems which might be investigated by the electro-optical method, as for instance, the edge corrections of parallel plate condensers and the distortion of a homogeneous field by obstacles of various shapes. Inasmuch as these photographs serve to find the solution of complicated boundary value problems involving Laplace's equation, the method is apt to furnish valuable information for problems in other fields where Laplace's equation is of importance, e.g., in hydrodynamics.

The experimental method can be improved in various ways. The conductivity of the colloids can be reduced by dialysis. Stroboscopic illumination or the use of fields of higher frequencies will improve the sharpness of the isochromatic lines. By employing a longer optical path and color photography it will be possible to determine from a single picture the field intensity in any point within a large area of the field. For accurate measurements it would be advisable to mount the Kerr cell in a thermostat and to record the intensity variations by microphotometric methods.

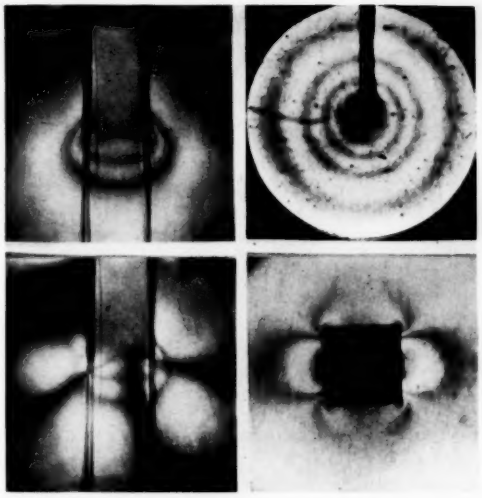


FIG. 4. Electro-optical field maps. At left: isochromatic lines and the isoclinic lines for 22.5° of the field at the edge of a parallel plate condenser. At right: isochromatic lines for the field in a cylindrical condenser and the distortion by a rectangular insulator of a uniform field between two parallel vertical plates.

The development of this new method would have been impossible without the collaboration of Professor E. A. Hauser and Dr. D. S. le Beau and their co-workers who prepared the bentonite gels, and of Dr. B. W. Sakmann, whose investigations clarified the electro-optical phenomena in the colloidal solutions. It is a pleasure to acknowledge the continued interest and assistance of these co-workers and to thank Mr. G. J. Yevick for technical assistance.

A New Quarter-Wave Compensator

M. RICHARTZ
Fu-Jen University, Peking, China
(Received November 20, 1940)

A modified quarter-wave compensator is described. The chief feature of the new arrangement is the "variable" half-shade plate whereby the accuracy of measuring very small phase-differences is increased. The new compensator is especially useful to calibrate compensators of the "variable order" type.

OPTICAL compensators[1] are used to measure the phase-difference of doubly refracting media by compensating devices. According to the way in which the phase-difference is measured, we may distinguish three general types of compensators: (1) the directly compensating method; (2) the compensator-azimuth measuring device; (3) the analyzer-azimuth measuring type. In the following we shall restrict our discussion to measurements of very weak double refraction and hence to compensators with half-shade devices.

The directly compensating method compensates the phase-difference in question by means of a doubly refracting plate of known phase-change. To this kind belong the Babinet-Soleil compensator[2] and the DeForest Palmer compensator.[3] The former makes use of the fact that in a doubly refracting, plane-parallel plate the path-difference of double refraction is changed by varying the thickness of the plate. The difference in the thickness gives the phase-difference to be measured. A thin, doubly refracting (mica) plate, which covers half the field only, serves as the half-shade plate. The Palmer compensator, the idea of which will be applied in the new compensator, produces double refraction of variable magnitude by applying a tensile stress to a glass plate. A half-shade strip of strain-free glass as well as a compensating strain-free glass plate can be strained by steel spiral springs; the tension in either case is regulated by micrometer screws.

In the second kind, the compensator-azimuth measuring device, the relation between the measured magnitudes and the phase-difference in question is more complicated than in the first one. Typical forms of this kind are the well-known Brace compensator[4] and the so-called "simple half-shade" compensator,[5] described by G. Szivessy. While in Brace's method the half-shade crystal-plate is fixed and the compensating crystal plate is rotated, Szivessy uses the half-shade plate itself as compensating device. Both methods calculate the phase-difference to be determined from the half-shade azimuths of the compensating plate.

The first and the second kind of compensators introduce the doubly refracting medium and the compensating device between the fixed positions of "crossed" polarizers; the third kind, however, turns the analyzer and thus by finding its half-shade azimuths determines the phase-difference of the medium. Here three compensators may be mentioned: two of these are the "modified" Brace and Szivessy compensators,[6] the third is the quarter-wave compensator.[7] The phase-difference measurable with the original Brace and

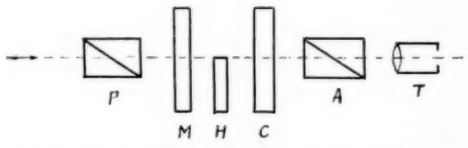


FIG. 1. Arrangement of the optical system: *P*, polarizer; *M*, medium to be measured; *H*, half-shade plate; *C*, quarter-wave plate; *A*, analyzer; *T*, telescope.

[1] G. Szivessy, *Handbuch der Physik*, Vol. 19, p. 935 (1928). Other references are given in the same article.
[2] A. Bravais, Comptes rendus **32**, 115 (1851); G. Szivessy, Zeits. f. Physik **29**, 372 (1924).
[3] DeForest Palmer, Phys. Rev. **17**, 409 (1921); H. A. Boorse, Phys. Rev. **46**, 187 (1934).
[4] D. B. Brace, Phys. Rev. **18**, 70 (1904); Wedeneewa, Zeits. f. Instrumentenk. **43**, 17 (1923).
[5] G. Szivessy, Zeits. f. Krist. **77**, 239 (1931); G. Szivessy and A. Dierkesmann, Ann. d. Physik [5] **11**, 949 (1931).
[6] G. Szivessy and W. Herzog, Zeits. f. Instrumentenk. **57**, 305 (1937).
[7] MacCullagh, Proc. Roy. Irish Acad. **1**, 158 (1841); De Sénarmont, Ann. chim. phys. (2) **73**, 337 (1840); F. Gabler and P. Sokob, Zeits. f. Instrumentenk. **58**, 301 (1938).

Szivessy compensators is of limited range. Szivessy and Herzog, however, changed these methods in such a way that not the azimuths of the compensating plate are measured, but the analyzer-azimuths while the compensator is left in fixed position, so that any phase-difference may be measured without restriction. If the special value $\lambda/4$ is given to the phase-difference of the compensating plate in the modified Brace compensator, we have the so-called "modified" quarter-wave compensator.[8]

The optical compensators may be viewed in another way. In all the methods mentioned above, except the quarter-wave compensators, two different *elliptical* radiations are to be analyzed, while in all quarter-wave compensators the light entering the analyzer is *linearly* polarized. But the modified quarter-wave compensator differs from the other half-shade quarter-wave compensators insofar as there are already two linear radiations emerging from the quarter-wave plate with different directions of vibration. The half-shade plate influences the elliptical radiation before it enters the quarter-wave plate. In all other quarter-wave compensators one linear radiation emerges from the quarter-wave plate and then it is divided into two parts of different directions of vibration by a half-shade device.

The new quarter-wave compensator is an application of Palmer's idea to the modified quarter-wave compensator. The arrangement is as follows: Between the Nicols P and A are placed the doubly refracting medium M of the phase-difference δ_1 to be measured, the half-shade plate H of variable phase-difference δ_2 and the compensating plate C of phase-difference $\delta_3=\lambda/4$ (Fig. 1). They are adjusted in such a way that when a monochromatic, parallel beam of light passes through the polarizer P, the emerging linear radiation falls normally upon the three plates M, H, and C. We relate the "fast" directions of vibration within the plates, p_1, p_2, p_3, to the direction of vibration p of the polarizer. Let a be the orientation of the direction of vibration in the analyzer A, referred to P (Fig. 2). The values of the corresponding azimuths φ_1, φ_2, φ_3 are given by

$$\varphi_1=\pi/4; \quad \varphi_2=\pm\pi/4; \quad \varphi_3=\pi/2 \text{ or } 0. \quad (1)$$

[8] M. Richartz, Zeits. f. Instrumentenk. **60**, 357 (1940).

The analyzer-azimuths $\bar{\alpha}$ for the half-shade positions are measured. It is sufficient to consider only one of the four analyzer positions. According to our restriction of measuring very weak double refraction let the phase-differences δ_1 and δ_2 be limited as follows:

$$0<\delta_1<\pi/2; \quad 0<\delta_2<\pi/2.$$

Let the linearly polarized beam of light pass through the doubly refracting plates M and H.

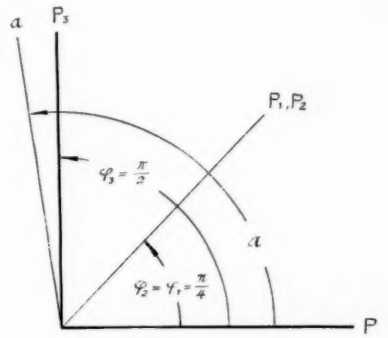


FIG. 2. Directions of vibration and their azimuths.

The emerging resultant vibration is elliptically polarized; the axes of the ellipse are parallel and normal to the direction of vibration of the polarizer.[9] The components of the ellipse with respect to these axes have a mutual phase-difference of $\pi/2$. Then the elliptical radiation passes through the quarter-wave plate C whose directions of vibration are parallel to the axes of the ellipse; therefore a phase-difference of $\pi/2$ is added. The resulting vibration is linearly polarized; its direction depends upon the ratio of the axes of the ellipse, hence upon the phase-differences of the doubly refracting plates.

When looking through the telescope T, focused to the half-shade plate, we see, in general, the field of view consisting of two parts of unequal brightness because of the unequal phase-differences in the two halves. One half of the radiation passes through both plates, while the other half is not influenced by the half-shade plate.

For the field of view covered with the plate M only the intensity J_1 of light passing through the

[9] G. Szivessy, *Handbuch der Physik* Vol. 19, p. 919 (1928).

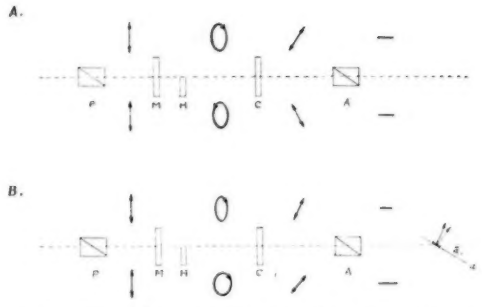


FIG. 3. Ellipticities of light vibrations caused by the different plates when $\delta_2=2\delta_1$. A—The "fast" axis of H is normal to that of M. B—The "fast" axis of H is parallel to that of M.

analyzer is given by

$$J_1=J_0\cos^2\left(\alpha\pm\frac{\delta_1}{2}\right), \tag{2}$$

where J_0 is the intensity of light emerging from the polarizer. The positive sign of δ_1 belongs to $\varphi_3=\pi/2$, the negative is valid when $\varphi_3=0$.

The intensity J_2 of the field of view covered with the two plates M and H is

$$J_2=J_0\cos^2\left(\alpha\pm\frac{\delta}{2}\right), \tag{3}$$

where $\delta=\delta_1\pm\delta_2$; the positive sign of δ_2 is valid for $\varphi_1=\pi/4$; $\varphi_2=\pi/4$; the negative for $\varphi_1=\pi/4$; $\varphi_2=-\pi/4$.

The intensity $J=J_1=J_2$ for a matched field gives the relation between the half-shade azimuths $\bar{\alpha}$ of the analyzer and the phase-differences δ_1 and δ_2 as:

$$2\bar{\alpha}=\pi\mp\left(\frac{\delta_1}{2}+\frac{\delta}{2}\right)$$

or

$$2\bar{\alpha}=\pi\mp\left(\delta_1\pm\frac{\delta_2}{2}\right). \tag{4}$$

According to Palmer's idea we use a glass slip as half-shade plate H. Since glass under tension behaves as a positive uniaxial crystal with the optic axis in the direction of the applied stress, our half-shade plate is equivalent to a doubly refracting plate of variable thickness. As long as there is no strain in the glass slip, the field of view is of equal brightness because of the equal intensities J_1 and $J_2(\delta_2=0)$. When the half-shade plate is strained, it produces an ellipticity in the opposite sense as that of the medium if the azimuths $\varphi_1=\pi/4$ and $\varphi_2=-\pi/4$. Consequently the part of the field which is covered with the glass slip becomes darker until for $\delta_2=\delta_1$ the ellipticity caused by the medium M is compensated by the phase-difference of the half-shade plate; for crossed Nicols the intensity J_2 is zero.

By increasing the tension the field becomes bright again and the match of the half-shade field will be restored when

$$\delta_2=2\delta_1. \tag{5}$$

The half-shade azimuth of the analyzer is

$$\bar{\alpha}_0=\pi/2. \tag{6}$$

After rotating the half-shade plate through 90° its azimuth φ_2 is $+\pi/4$ and δ is the sum of the phase-differences δ_1 and δ_2 *viz.* $3\delta_1$. Hence the half-shade azimuth

$$\bar{\alpha}_1=\pi/2\mp\delta_1 \tag{7}$$

gives us directly the phase-difference in question. The phase-change in the half-shade plate need not to be known. Figure 3 illustrates the given theory.

The procedure of measuring the phase-difference of a doubly refracting medium is, therefore, as follows: Between crossed polarizers the medium, the half-shade glass strip and the quarter-wave plate are arranged as described. The half-shade plate is adjusted in such a way that its azimuth φ_2 is opposite in sign to φ_1. Let the analyzer reading for its azimuth $\alpha=\bar{\alpha}_0=\pi/2$ be R_0. Then the half-shade plate is strained until there appears half-shadow. The stressed plate (or the medium) is rotated through 90°. Finally the analyzer is turned to find the half-shade position again; let this analyzer reading be R_1. The difference of the readings of the two analyzer positions gives the phase-difference δ_1 in degrees:

$$R_1-R_0=\delta_1. \tag{8}$$

To check the adjustment we now give to the quarter-wave plate the azimuth $\varphi_3=0$ by rotating it through 90° and take the analyzer readings while the half-shade plate remains stressed. If δ_2 is exactly equal to $2\delta_1$, then the reading R_0' is the same as R_0 and R_1' lies on the

opposite side of R_0 at the same angular distance as R_1. The difference again must be δ_1.

The sensitivity of the half-shade positions is given by the expression[10]

$$h=\frac{\left|\frac{\partial}{\partial\alpha}(J_2-J_1)\right|_{\alpha=\bar{\alpha}}}{\bar{J}},$$

where the intensities J_1 and J_2 are given by Eqs. (2) and (3) and the special values $\bar{J}$ and $\bar{\alpha}$ by Eq. (4).

We obtain for both analyzer readings R_0 and R_1 the value

$$h=4\ \text{cotan}\ \frac{\delta_1}{2}. \qquad (9)$$

It appears that the sensitivity increases with smaller values of δ_1. Therefore the new method is especially useful for the measurement of very weak double refraction.

The lowest limit depends among other conditions especially upon the divided circle of the rotating analyzer support. If the smallest graduated arc is $\frac{1}{6}°$ with a vernier for a reading of 10″, we get an accuracy of about $8\times10^{-6}\times2\pi$ for the phase-difference δ_1.

In all other half-shade devices of the quarter-wave compensator the difference of the two analyzer readings gives half the value of the phase-difference in question. Therefore the error in the final result is twice as great as the error of the measured value, while in the new method the error in δ_1 corresponds directly to the measured azimuth.

Finally *a special advantage* may be named, *viz.* the rather suitable manner of calibrating the Palmer or similar compensators. Introducing the "compensating" plate into the optical system as M plate the "calibration" curve of the compensating micrometer screw is very easily carried out because the analyzer reading gives directly the value of the phase-difference. It is only necessary that the support of the compensating plate be constructed in such a way that the plate can be turned in its plane through 90°.

Similarly Palmer's half-shade plate can be calibrated. If there is no doubly refracting medium M (no strain in the compensating plate), but the half-shade plate between the polarizer and the quarter-wave plate, the condition for a matched field is given by Eq. (4) when $\delta_1=0$:

$$\bar{\alpha}=\pi/2\mp\delta_2/4.$$

Using the quarter-wave plate in both azimuths $\pi/2$ and 0, we get with greater accuracy from the two readings R_1 and R_1' the phase-difference δ_2:

$$2(R_1'-R_1)=\delta_2.$$

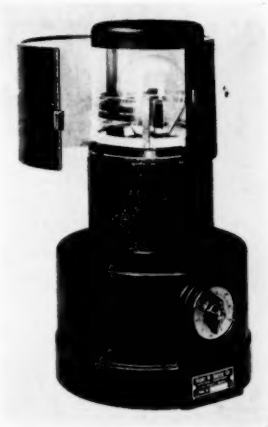
[10] G. Szivessy, *Handbuch der Physik* Vol. 19, p. 926 (1928).

Infra-Red Ray Film Dryer

THE Harry W. Dietert Company of Detroit recently announced a new Infra-Red Ray Film Dryer. The Dryer was developed for the rapid drying of spectrograms, but other films up to 18 inches in length may be dried with the unit.

The Infra-Red Ray Dryer consists of a cast aluminum base containing a blower, and an air filter to assure dust free air. The blown air is warmed by a Nichrome wire heating element and passes over the film, which is clipped around a Pyrex glass cylinder. A 375-watt infra-red lamp within the glass cylinder provides the heat for drying.

Depending on the type of emulsion, from one to four minutes are required to dry films. The drying time is regulated by an electric timer which shuts the drying elements off after the elapse of a predetermined interval.



The Photon Theory of Optical Resolving Power

B. P. Ramsay, E. L. Cleveland and O. T. Koppius
Department of Physics, University of Kentucky, Lexington, Kentucky
(Received December 31, 1940)

I. The Epoch-Slope Definition

In a previous paper,[1] it has been proposed that spectroscopic resolving power be defined by the formula

$$P=(\alpha/Y)(\Delta I/\Delta\alpha). \tag{1}$$

In this formula, α is a quantity which has been called the epoch function.[2] For a number of familiar cases, it is directly proportional to the order of interference. The ratio, $\Delta I/\Delta\alpha$, is the intensity-epoch slope. It was introduced to describe the dependence of resolving power upon the width and form of the intensity distribution produced by an optical system. This ratio is intimately related to criteria for the limit of resolution; a criterion is an essential part of any theory of resolving power.[3] The inverse factor, Y, is obviously necessary in the epoch-slope definition (1) to make resolving power a dimensionless quantity. There are reasons for believing, however, that resolving power is an inverse function of the magnitude of the intensity observed with a given optical system. It has been proposed, therefore, that Y be defined as the function which describes this dependence. In the present paper, arguments for the existence of this dependence, and its cause, will be presented from the point of view of the photon theory.

The procedure which will be followed here has the merit of further clarifying the relation between the epoch-slope definition (1) and the Rayleigh definition of spectroscopic resolving power. The Rayleigh definition of resolving power is

$$\rho=\lambda/\Delta\lambda=-\nu/\Delta\nu. \tag{2}$$

The quantum energy formula,

$$\epsilon=h\nu, \tag{3}$$

and the Heisenberg indeterminacy principle stated in the form

$$\Delta\epsilon\Delta t=h, \tag{4}$$

provide the relations which are necessary to translate the Rayleigh resolving power, ρ, into the idiom of the photon theory. In these relations, ϵ is the energy of a photon, ν is the corresponding frequency, and h is Planck's constant. $\Delta\epsilon$ is the uncertainty in the energy of a photon, and Δt is the uncertainty in its time specification. Expressed in terms of Δt, (2) becomes[4]

$$\rho=\nu\Delta t. \tag{5}$$

In this manner, the Rayleigh resolving power has been interpreted on the photon theory as the product of frequency by the time uncertainty of a single photon. One of the purposes of the present paper, therefore, is to show the relation between the Rayleigh resolving power, as given by (5), and the epoch-slope definition (1).

II. Observability and the Principle of Uncertainty

According to the photon theory, a beam of light consists of a large number of photons, all of which are alike in having zero charge and zero rest mass. The intensity of the beam is equal to the energy transmitted by the photons in the beam through unit area normal to their path in unit time. Let N be the number of photons which pass through a given normal unit area in unit time. Then, if the energy of each photon is given by an expression of the form (3) with a characteristic subscript, the intensity of the

[1] B. P. Ramsay, O. T. Koppius and E. L. Cleveland, "Physical interpretations of resolving power," J. Opt. Soc. Am. **31**, 202 (1941).
[2] B. P. Ramsay, O. T. Koppius and E. L. Cleveland, J. Opt. Soc. Am. **30**, 439–444 (1940).
[3] B. P. Ramsay, O. T. Koppius and E. L. Cleveland, "Criteria and the intensity-epoch slope," J. Opt. Soc. Am. **31**, 26 (1941).

[4] Algebraic manipulation may suggest that a negative sign should properly appear on the right side of (5). However, the sign has been completely absorbed in Δt. The correct method of handling the sign has been discussed in a previous paper; see, for example, reference 1.

beam is

$$E=\sum_{j=1}^{N}\epsilon_j=\sum_{j=1}^{N}h\nu_j. \tag{6}$$

However, if the beam is monochromatic, such that for every value of j,

$$\nu_j=\nu, \tag{7}$$

the intensity (6) becomes

$$E=N\epsilon=Nh\nu. \tag{8}$$

These relations, (7) and (8), are approximately applicable to observations made upon a single spectrum line. When employed for this purpose, however, they are not to be interpreted as denying the possibility of a frequency distribution in the line, but rather that the frequency distribution of *observable* photons is small. Actually, the frequency distribution may be indefinitely large; but the density of photons at any large frequency interval from the center of gravity of the given spectrum line is so small as to be unobservable. In this case, the value of ν which appears in (7) and (8) is the frequency of the center of gravity of the spectrum line.

This point of view suggests a first deficiency of the Rayleigh resolving power, which is expressible either in terms of $\Delta\epsilon$, the uncertainty in the energy of a single photon, or in terms of Δt, the uncertainty in its time specification. The Heisenberg uncertainty relation holds for each photon, but is not immediately applicable to experimental measurements for the reason that individual photons cannot be observed by laboratory procedures. In order to make use of it in discussing optical resolution, it is necessary to have an auxiliary relation between the uncertainty of photons which are present in sufficient density to be observed and the uncertainty of a single photon. To obtain such a relation, let ΔE be the uncertainty in energy of *observable* photons. Perhaps the simplest function which may be assumed for the auxiliary relation is

$$\Delta E=qN\Delta\epsilon, \tag{9}$$

where q is that fraction of the N photons which is observable. Because of the simplicity of (9), it will be employed first; other possibilities will be considered later.

There must be some property of a spectroscopic observation which is determined by ΔE. Let this property be designated by r. From (2) and (3), it follows that the Rayleigh resolving power may be written as

$$\rho=h\nu/\Delta\epsilon=\epsilon/\Delta\epsilon. \tag{10}$$

Thus, by analogy, the undefined property r may be written as

$$r=g/\Delta E, \tag{11}$$

where g is the factor of proportionality. The use of (4) in (9) gives

$$\Delta E\Delta t=qN\Delta\epsilon\Delta t=qNh, \tag{12}$$

or

$$\Delta E=qNh/\Delta t. \tag{13}$$

Substitution of (13) in (11) yields

$$r=g\Delta t/qNh. \tag{14}$$

Equations (11) and (14) are equivalent, but (14) is more serviceable in obtaining an expression for r in terms of properties of the spectroscope. It will first be used to discuss a grating; afterwards, its application to a prism will be noted.

III. Prismatic and Multiple Beam Devices

In (14), Δt is the uncertainty in time specification of a single photon which arrives at a given point in the focal plane of a spectroscope. This uncertainty in time arises from the lengths or the nature of the paths traversed by the photons. In the case of gratings and other multiple beam arrangements, it results from the inequality in the lengths of available optical paths which the photon may have traversed in passing through the spectroscope. The time uncertainty is introduced by prismatic devices in a different way, for the paths taken by photons in traveling from the source to the point of observation are optically equal according to Fermat's principle. Photons travel these paths with energy velocities (group velocities) which are a function of the medium. Thus an uncertainty in time is introduced by a prismatic spectroscope only because photons travel with different energy velocities in different media.

The time uncertainties introduced by prisms and gratings have been calculated by Goldman.[5]

[5] S. Goldman, J. Opt. Soc. Am. **23**, 70–71 (1933).

The calculation has been made[1] for multiple reflection devices also—such as the Fabry-Perot interferometer. For a grating, it is

$$\Delta t = km\lambda/c, \tag{15}$$

where k is the number of beams, m is the order of the spectrum, λ is the wave-length, and c is the velocity of light. When (15) is substituted in (14), it appears that

$$r = (g/qNh)(km\lambda/c), \tag{16}$$

or

$$r = gkm/qNh\nu = gkm/Y, \tag{17}$$

where, by the use of (8),

$$Y = qNh\nu = qE, \tag{18}$$

and represents that portion of the energy in the intensity distribution of a spectral line which is observable. It is to be noted that this result is applicable not only to gratings but to all multiple beam arrangements.

For a prism,

$$\Delta t = (T/v)(d\mu/d\lambda), \tag{19}$$

where T is the thickness of the prism and μ is its refractive index. For this case, (14) becomes

$$r = (g/qNh)(T/v)(d\mu/d\lambda), \tag{20}$$

or

$$r = (g/Y)(Td\mu/d\lambda). \tag{21}$$

By (18), it appears that the quantity r, in (17) and (21), is inversely proportional to the intensity, E. However, reconsideration shows that this result was written into the derivations by (9) and (11). A change either in the definition of r, or in the auxiliary relation between ΔE and $\Delta\epsilon$ would have led to a different result. Nothing is to be gained, it may be noted, by changing the definition of r; for the whole point of the inquiry is to determine the characteristics of a quantity, r, defined analogously to ρ with the substitution of ΔE for $\Delta\epsilon$. On the basis of (9) and (11), therefore, it may be concluded that r is different from ρ unless

$$g = qNh\nu = qE; \tag{22}$$

for only so do (17) and (21), respectively, reduce to

$$r = km = \rho \tag{23}$$

for multiple beam arrangements, and to

$$r = Td\mu/d\lambda = \rho \tag{24}$$

for prismatic devices. On the other hand, if g is a constant of the apparatus which does not involve the intensity, r must be inversely proportional to E.

It is still possible, of course, that the relation between ΔE and $\Delta\epsilon$ may not have the form assumed in (9). Nevertheless, it appears that the dependence (or independence) of r on E, expressed by the function Y, is entirely governed by the relation between ΔE and $\Delta\epsilon$. The nature of Y, therefore, is to be determined empirically. Experimental studies of the Y function in visual and photographic observations will be discussed in a subsequent paper.

IV. The Definition of Resolving Power

In the preceding discussion, ΔE is the total uncertainty in the energy of the photons which are observed, but $\Delta\epsilon$ is the theoretical uncertainty in the energy of a single photon. The source, the spectroscope, and the observing mechanism all contribute to the uncertainty, ΔE. On the other hand, $\Delta\epsilon$ is only that portion of the uncertainty which is introduced by the spectroscope itself. Thus, r is a measure of the actual resolution which may be effected by an optical system; but it is not suitable for comparing prisms, gratings, and other isolated spectroscopic devices unless the entire system and procedure be specified. On the other hand, ρ offers a purely theoretical basis for comparing optical instruments, but the rating of the instruments thus obtained is not a measure of the resolution effected in any particular case.

It has already been pointed out that ρ as given by (10) is the limiting theoretical resolving power of an isolated optical device as given by the Rayleigh formula, $\rho = \lambda/\Delta\lambda$. The use of hypothetical measurements of idealized simplicity as a basis for predicting results which one may expect to obtain from laboratory experiments has been considered erroneous.[6] *In the case at hand, it appears neither desirable nor necessary to identify resolving power with ρ, as*

[6] R. B. Lindsay, Am. J. Phys. **8**, 22–27 (1940).

given by the usual Rayleigh formula. The alternative proposed is that resolving power, P, be identified with r, the resolution which can be effected by an optical system. This identification is consistent with experience in two respects: (A) It is known that increasing the intensity of a source is likely to be accompanied by increased thermal agitation of the molecules of the source. It is accompanied also by intra-atomic and intramolecular disturbances. These effects are classically related to problems of resolution by means of collision damping and the Doppler effect. (B) Moreover, resolving power is affected by the procedure which is employed in making visual or photographic observations. The characteristics of the photographic plate and its development, as well as the sensitivity of the microphotometer and the manner in which it is used, are significant factors in determining what resolving power may be obtained. These factors have been clearly recognized, and studied in detail in photographic laboratories.[7]

Thus, although the identification,

$$P=r=g/\Delta E, \tag{25}$$

differs from the Rayleigh formula, as has been noted, it has the advantage of being applicable to practical laboratory measurements, and is consistent with the conceptual significance which long usage has attributed to resolving power; *viz.*, the greatest resolution which can be attained with a given optical system. Moreover, the identification (25) is entirely compatible with the epoch-slope definition of resolving power given by (1). Although the definition (1) was proposed on the basis of arguments of a different type, the present approach from the point of view of the photon theory is valuable in that it indicates the role of the function Y in this definition. The previous consideration merely required that Y serve the comparatively trivial purpose of adjusting units in such a way as to make P a dimensionless quantity. It now appears, however, that Y is inherently related to the threshold of observability, and must describe the significance of this threshold with respect to the resolution obtainable with a given optical system.

[7] F. E. Ross, *The Physics of the Developed Photographic Image* (D. Van Nostrand Company, 1924), Chapter IV.

V. The Identification of r with P

It will be recognized that the formula for resolving power given by (25) does not display all the quantities upon which P depends. For instance, the effect of the shape and width of a spectrum line, which appears in (1) as the intensity-epoch slope, is not expressly presented in (25). Moreover, it is common knowledge that resolving power is a function of the instruments used for observation. It also depends upon the quality of the workmanship and materials used in the construction of the spectroscope. Nevertheless, (25) is consistent with these facts, as can be seen by comparing (1) and (25) in detail.

In order to avoid misunderstanding, it is perhaps desirable to emphasize the fact that the value which is assigned to g in (22) in order to obtain the Rayleigh formula is not the value which is to be given g in the more inclusive formula (25). In (25), g is the criterion coefficient which was introduced in the previous paper[1] to show the relation between ρ and P.

The intensity-epoch slope, $\Delta I/\Delta\alpha$, of (1) is included in (25) through ΔE. If the prism of a spectroscope contains air bubbles, or if the rulings of a grating are defective, the uncertainty ΔE increases, whereas $\Delta I/\Delta\alpha$ decreases in such a way as to make the values of P given by (1) and (25) equal. In the case of optical defects, ΔE increases because $\Delta\epsilon$, the uncertainty of each photon, increases. The magnitude of the effect upon ΔE is determined by the auxiliary relation between ΔE and $\Delta\epsilon$.

The value of q in (9) depends upon such quantities as the properties of the photographic process, the candlepower of the microphotometer lamp, the reaction of the photo-cell, the sensitivity of the galvanometer, and other data concerning procedure of observation. Any process which reduces the value of q has the effect of increasing the resolving power, P. There are, for instance, procedures for increasing the contrast obtained in a given negative by transferring the photographic negative to a positive, the positive to a negative again, and so on repeatedly. However, in this process, a value of gamma different from unity must be used for the D log E sensitometric curve. For if unit value of gamma were used, the only effect of each transfer

would be to make blacks white, and whites black without changing the contrast in any quantitative way. In this case, identical conclusions concerning the resolution effected could be drawn from either positive or negative.

In regard to the effect of diminishing q, it is interesting to consider the absorbing prism.[2] For it appears that absorption in the prism should have the effect of increasing P by diminishing q. The absorption of the prism, however, has the further effect of increasing the uncertainty $\Delta\epsilon$—or of decreasing the uncertainty Δt as given by (19)—to such an extent that the net result is a decrease of resolving power with the absorption of the prism.

Finally, it may be remarked that the inverse dependence of resolving power on intensity, which has been noted in (17) and (21), is not to be applied to processes of observation which do not involve considerations of resolution. Thus, in the use of the ultra-microscope, it is known[8] that the "visibility" of submicroscopic particles increases with intensity. In this case, however, light from the source does not pass directly into the microscope, and there is no attempt to "resolve" the particles in such a way as to determine their size, shape, or structure.[9] Observations are made only to determine relative size, or relative concentration. Increasing the intensity of the source merely serves to bring the luminous energy scattered by smaller and smaller particles above the threshold of observability. It is true also in the use of a microphotometer that a high illumination of the photometric field gives improved discrimination over low brightness. The candle power of the microphotometer standard lamp affects the resolving power through the value of q. It is obviously not the intensity of this lamp to which the quantity, E, refers.

ACKNOWLEDGMENT

It is a pleasure for the authors to acknowledge the use of suggestions offered by Dr. F. L. Yost.

[8] K. B. Merling-Eisenberg, J. Quekett Microscopical Club **1**, 311–323 (1937); see, for example, F. Jaentzsch, *Handbuch der Mikrobiol. Tecknik*, Vol. 1, 49 (1921).

[9] O. E. Meyer, *The Diffraction of Light, X-rays, and Material Particles* (University of Chicago Press, 1934), p. 211.

Ninth Summer Conference on Spectroscopy and Its Applications

A NINTH Summer Conference on Spectroscopy and Its Applications will be held at the Massachusetts Institute of Technology on July 21, 22 and 23 of this year. Morning and afternoon sessions will be held during three days, with discussions by qualified experts on spectroscopic analysis of materials and on other applications of spectroscopy to biology, medicine, chemistry, metallurgy and mineralogy, and to industrial and engineering problems.

This conference will repeat to a considerable extent the type of program presented at the conferences held during the past eight summers. The early sessions will be devoted principally to applications of spectroscopy to the analysis of materials. Other topics to be considered are absorption spectrophotometry, photographic photometry, biological and chemical effects of spectral radiation, industrial applications of spectroscopy, analysis of spectra and the measurement of wave-lengths. The more technical and theoretical papers will be reserved for the later sessions.

As soon as it has been prepared, copies of the detailed program of the conference will be sent to anyone interested. Attendance at the conference will be limited to 200, and experience in previous years indicates that those who plan to attend from a distance should signify their intention in advance, to insure availability of space. No fees are charged for attendance. The object of the conference is to promote cooperation among workers in various scientific and industrial fields who have found useful the technical methods of spectroscopy.

The Proceedings of the Sixth and Seventh Conferences have been published and are still available. The Proceedings of the Eighth Conference have appeared in the January, February and March, 1941 issues of this Journal.

Interreflections in Lightwells

PARRY MOON
Electrical Engineering Department, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received December 10, 1940)

1. INTRODUCTION

IN many problems of natural or of artificial lighting, multiple reflections of light are not of primary importance and can be neglected. In other cases, such as the luminosity distribution in lightwells with walls of high reflection factor, interreflections must be considered if the calculated values are to be even a rough approximation to the true values.

Thanks to the pioneer work of Buckley,[1] the method for the solution of the lightwell problem by means of integral equations is well established. Hisano[2] has also contributed to the subject, while W. T. White[3] has recently obtained some interesting solutions by means of the cinema integraph. Others[4,5] have chosen to dispense with integral equations and by an exceedingly laborious method of successive approximations have obtained numerical results in a few cases. Experimental investigations on models and on actual lightwells have also been made, notably by Meacock and Lambert.[6]

Thus the subject of the luminosity distribution in lightwells has been studied in considerable detail, and methods of calculation have been established. Nevertheless, it is felt that another contribution to the subject will be of value in simplifying the results and in presenting a unified picture of the whole field. The equations and curves of the present paper are a direct application of the general formulas developed in a preceding paper[7] and are based on the representation of the kernel by a single exponential function.

2. THE KERNELS

Consider a rectangular lightwell having a cross section $1\times w$ and a depth l. The top of the well is open to a uniform sky of luminosity L_u. Obviously, a number of assumptions must be made to get the actual physical problem into a form simple enough for mathematical analysis. In the first place, the actual lightwell is far from uniform in reflection factor because of the windows. This condition will be idealized by assuming uniform walls having a reflection factor ρ_1, which is equal to the average reflection factor of the actual lightwell. The surfaces are considered to be perfectly diffusing. The assumption is also made that the luminosity is independent of the distance around the walls but is a function of the depth. This condition has been found to hold reasonably well,[6] except near the top of the lightwell.

At any distance s below the top edge of the lightwell, the average luminosity of the walls is given by an integral equation of the Fredholm type having the kernel:[7]

$$K(s-t)=\frac{1}{\pi(1+w)}\left\{ln\left[\frac{w^2+(s-t)^2}{(s-t)^2}\right]-ln\left[\frac{1+w^2+(s-t)^2}{1+(s-t)^2}\right]+w\left[\frac{1}{[1+(s-t)^2]^{\frac{1}{2}}}\tan^{-1}\frac{w}{[1+(s-t)^2]^{\frac{1}{2}}}+\frac{w}{[w^2+(s-t)^2]^{\frac{1}{2}}}\tan^{-1}\frac{1}{[w^2+(s-t)^2]^{\frac{1}{2}}}\right]\right\}. \quad (1)$$

Equation (1) was used in plotting the kernels for $w=1$, 2, 4, 8, and ∞ (Fig. 1). Between the abscissae 0 and approximately 0.3, the curves are distorted progressively as w is changed; but for larger values of $|s-t|$, all curves appear to be

[1] H. Buckley, "The blackness of blackbodies and the illumination of lightwells," J. Opt. Soc. Am. **18**, 216(1929).
[2] K. Hisano, "Fundamental problems on daylight lighting," Res. Elec. Lab., Tokyo, No. 397 (1936).
[3] W. T. White, "Calculation of the light distribution in lightwells," J. Opt. Soc. Am. **31**, 308 (1941).
[4] A. D. Moore, "Interreflections by the increment method as applied to a light court," I.E.S. Trans. **24**, 629 (1929).
[5] E. W. Manning and H. D. White, "Brightness distribution due to interreflection in a light court," I.E.S. Trans. **25**, 663 (1930).
[6] H. F. Meacock and G. E. V. Lambert, "Efficiency of lightwells," Dept. Sci. and Ind. Res., Tech. Paper No. 11 (1930). C.I.E. Proc. p. 1236 (1931).
[7] Parry Moon, "Interreflections in finite cylinders," J. Opt. Soc. Am. **31**, 223 (1941).

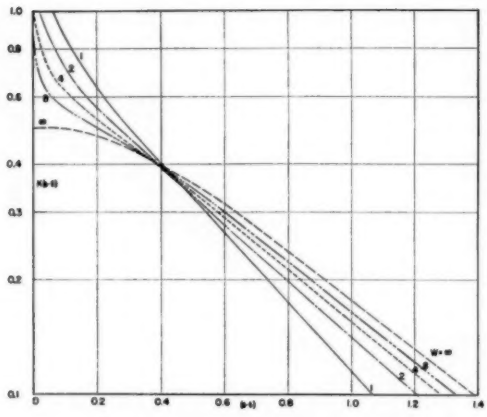


FIG. 1. Kernels for the rectangular parallelepiped.

nearly straight. Thus it is proposed to approximate all the kernels by straight lines, or by the equation

$$K(s-t) \cong Ae^{-a|s-t|}, \tag{2}$$

where the constants A and a are determined to give representations of the curves of Fig. 1.

All the actual kernels, except the one for $w=\infty$, go to ∞ at $(s-t)=0$. The trouble is caused by the first term of Eq. (1) and is found to be associated with the mathematically sharp corners of the lightwell. Though these infinities might conceivably bother the mathematician, they will not worry the engineer. He knows that the infinities can be eliminated by slightly rounding the corners of the lightwell or by painting the corners black, and his physical intuition tells him that neither of these changes will have any appreciable effect on the final luminosity distribution $L_1(s)$.

Another justification for neglecting the irregularities in the curves near $(s-t)=0$ was obtained by approximating the kernel for $w=1$ by the sum of two exponentials. One exponential, Eq. (2), took care of the part of the curve for $(s-t)>0.3$, while a highly damped term was used near zero. The sum of the two terms was a very good approximation to the actual kernel for $0.01<(s-t) \leqq 1.0$. Solution of the integral equation for a semi-infinite cylinder and with the two-term approximation gave the same result (within 2 percent) as that obtained with the one-term kernel. Somewhat similar conclusions were obtained by Yamauti[8] for the circular cylinder. One-term, two-term, and three-term approximations were used; and in general the one-term kernel gave *better* results than the two-term, besides being immensely simpler.

The next step is to decide on the values of A and a in Eq. (2). If the approximate kernel is to give correct results for the semi-infinite cylinder,[7]

$$A=a/2, \tag{3}$$

which reduces the number of parameters to one. It is necessary merely to determine a value of a corresponding to each value of w. This determination may be made graphically by use of Fig. 1. However, a simpler way is to utilize the fact[7] that for correct results for $l \to 0$,

$$a=p/2S, \tag{4}$$

where

p = perimeter of the cross section,
S = area of the cross section.

Equations (3) and (4) give the values of the parameters of Eq. (2) for any shape of cross section of the lightwell.

3. Solution of the Integral Equation

The solution[7] of the integral equation for the rectangular lightwell with nonreflecting bottom and with $\rho_1<1$ is

$$L_1(s)=B \cosh ks+C \sinh ks+D, \tag{5}$$

where

$$B=\rho_1 L_u \frac{(1-\rho_1)^{\frac{1}{2}} \cosh kl+\sinh kl}{2(1-\rho_1)^{\frac{1}{2}} \cosh kl+(2-\rho_1) \sinh kl}, \tag{6}$$

$$C=-\rho_1 L_u \frac{\cosh kl+(1-\rho_1)^{\frac{1}{2}} \sinh kl}{2(1-\rho_1)^{\frac{1}{2}} \cosh kl+(2-\rho_1) \sinh kl}, \tag{7}$$

$$D=0, \tag{8}$$

$$k=a(1-\rho_1)^{\frac{1}{2}}, \tag{9}$$

$L_1(s)$ = average luminosity of wall at depth s,
L_u = luminosity of the sky,
ρ_1 = average reflection factor of walls.

If the depth of the lightwell is sufficient so that $kl>1.5$, Eq. (5) reduces to the simple approxi-

[8] Z. Yamauti, "On the blackness of cylindrical blackbody," Res. Elec. Lab., Tokyo, No. 378 (1934).

mate formula:

$$L_1(s) = L_u[1-(1-\rho_1)^{\frac{1}{2}}]e^{-ks}. \qquad (10)$$

Thus, except in lightwells of small depth, the luminosity of the walls varies exponentially with the distance s from the top. The luminosity distributions for a number of lightwells are plotted in Fig. 2 for a constant reflection factor of 0.80. As might be expected, luminosity deep in the well is increased as the length w of the well is increased, but luminosity near the top of the well is almost independent of w. The case of $w=\infty$ is approximated by long, parallel rows of buildings on the two sides of a street.

Equation (10) was also used in plotting Fig. 3, which shows the effect of reflection factor. This figure was calculated for a square lightwell ($w=1$) but can be used equally well for any value of w, as was pointed out in a previous paper.[7] For $w=1$, the value of a is 2 and the abcissae represent actual distances s. For a lightwell of any other cross section, the proper value of a is obtained from Eq. (4) and actual distances are multiplied by $a/2$ to get abcissae. For a well of infinite length, for example, $a=1$ and the horizontal scale on the graph sheet represents $s/2$.

The luminosity distribution in a $1\times1\times4$ lightwell, having a nonreflecting bottom and a wall reflection factor of 0.80, is given by Manning and White.[5] They use a laborious method of numerical approximation which finally results in the points represented by circles in Fig. 4. The straight line represents the theoretical distribution according to Eq. (10). This simple equation gives values that are slightly higher than those of Manning and White at $s=0$ and somewhat lower at large values of s, but the agreement is sufficiently good for engineering purposes.

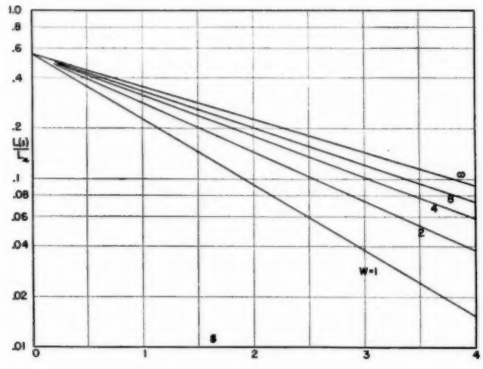


FIG. 2. Calculated luminosity distributions on the walls of lightwells of various lengths.
$\rho_1=0.80$, $l\to\infty$.

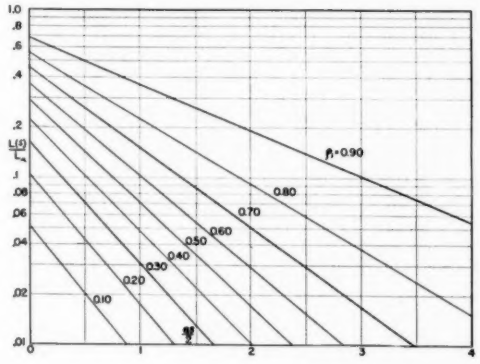


FIG. 3. Calculated luminosity distributions on the walls of a lightwell, showing how the result is affected by the reflection factor. $l\to\infty$.

Experimental results of Meacock and Lambert[6] are shown by the black circles in Fig. 4. The data were obtained by means of a Macbeth illuminometer and a model, $1\times1\times20$, coated with a matt paint having a reflection factor of 0.80. The results calculated by means of the simple formula, Eq. (10), lie approximately midway between the Manning-White calculated values and the Meacock-Lambert experimental values. Whether the discrepancies between the two latter sets of results is caused by experimental error (cosine-law error in the photometer) or is a true difference to be expected in lightwells is not evident. In either case, however, the use of the exponential kernel, Eq. (2), is vindicated where only ordinary engineering accuracy is required.

Figure 5 gives additional data for walls of various reflection factors. The points represent the experimental results of Meacock and Lambert, while the straight lines are obtained by use of Eq. (10). The agreement is very good for $0\leqq s\leqq2$; but at greater depths the calculated results are lower than the experimental ones. This discrepancy may well be caused by the approximate kernel or it may be an experimental error.

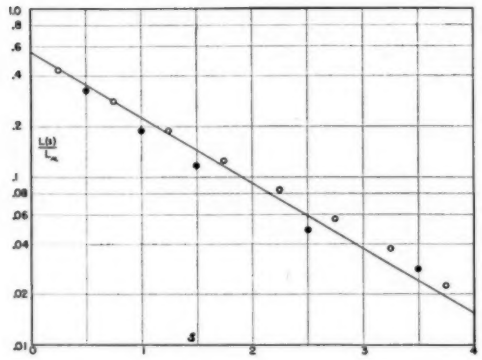


FIG. 4. Luminosity distribution for a square lightwell ($\rho_1=0.80$). ○ Manning and White; ● Meacock and Lambert; —— calculated ($l\to\infty$).

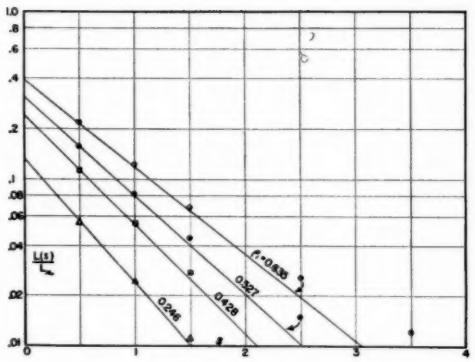


FIG. 5. Luminosity distribution for a square lightwell. ○ Manning and White; ● Meacock and Lambert; —— calculated ($l\to\infty$).

4. SOLUTION FOR $\rho_1=1.00$

Manning and White[5] give calculated luminosity distributions for reflection factors of 0.80 and of unity. For $\rho_1=1.00$, however, Eqs. (5) and (10) do not apply[7] and it is necessary to derive a new equation.

The integral equation[7] is

$$L_1(s)=L_{01}(s)+\rho_1\int_0^l K_{11}(s-t)L_1(t)dt$$

$$=\frac{L_u}{2}e^{-as}+\frac{a}{2}\left\{e^{-as}\int_0^s e^{at}L_1(t)dt\right.$$

$$\left.+e^{as}\int_s^l e^{-at}L(t)dt\right\}. \quad (11)$$

Assume a solution of the form

$$L_1(s)=A-Bs \quad (12)$$

and substitute in Eq. (11). The result is

$$A=L_u\left(\frac{al+1}{al+2}\right), \quad B=\frac{L_u a}{al+2}. \quad (13)$$

Thus the solution is

$$L_1(s)=\frac{L_u}{al+2}\{(al+1)-as\}. \quad (14)$$

A surprising feature of this equation is that at half the total depth ($s=l/2$), the value of $L_1(s)/L_u$ is always exactly $\frac{1}{2}$, irrespective of a.

Values calculated by Eq. (14) are compared in Fig. 6 with the values of Manning and White. The data are seen to be in good agreement. Both sets are representable by straight lines, but the approximation of Manning and White has a slightly smaller slope than that calculated from Eq. (14). W. T. White[3] has recently shown that the true curve is not exactly straight but has a slight departure from linearity. The true curve goes through the point ($s=l/2$, $L_1(s)/L_u=1/2$) as predicted by Eq. (14) and the departure is

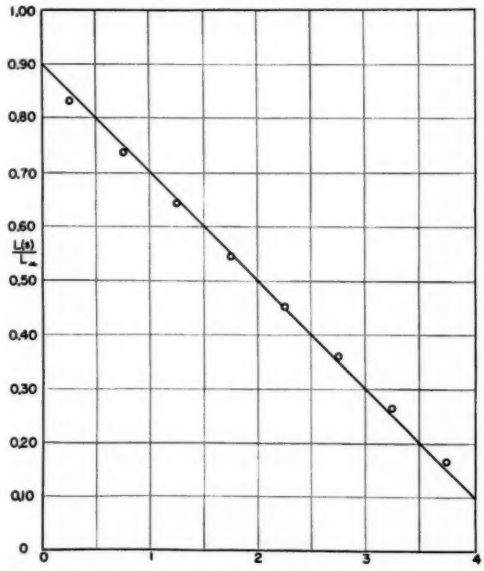


FIG. 6. Luminosity distribution for a square lightwell. ($\rho_1=1.00$, $\rho_3=0$, $l=4$.) ○ calculated by Manning and White; —— calculated by use of Eq. (14).

symmetric about the point. This slight correction is of theoretical interest but is of no importance for most engineering purposes.

5. Lightwell with Reflecting Bottom

If the lightwell has a reflecting bottom with reflection factor ρ_3, the equation[7] for the wall luminosity becomes

$$L_1(s)=\frac{L_u\rho_1}{W}[U\cosh ks-V\sinh ks]. \quad (15)$$

The luminosity of the bottom is

$$L_3(l)=\rho_3L_ue^{-al}\Big[1+\frac{U}{W}\{e^{al}[\cosh kl-(1-\rho_1)^{\frac{1}{2}}\sinh kl]-1\}-\frac{V}{W}\{e^{al}[\sinh kl-(1-\rho_1)^{\frac{1}{2}}\cosh kl]+(1-\rho_1)^{\frac{1}{2}}\}\Big], \quad (16)$$

where

$$U=(1-\rho_1)^{\frac{1}{2}}(1+\rho_3)\cosh kl+(1-\rho_3)\sinh kl,$$

$$V=(1-\rho_1)^{\frac{1}{2}}(1+\rho_3)\sinh kl+(1-\rho_3)\cosh kl,$$

$$W=2(1-\rho_1)^{\frac{1}{2}}\cosh kl+[(1-\rho_3)+(1-\rho_1)(1+\rho_3)]\sinh kl.$$

Equation (15) was used in computing the luminosity distribution in a square lightwell having a reflection factor of 0.80 for sides and bottom. The result is shown by the heavy curve of Fig. 7. The curve is in good agreement with the points, representing the values of Manning and White, from $s=0$ to $s=2$; but beyond this value, the curve is much too low.

The discrepancy noted in Fig. 7 is caused by the exponential approximation to the kernel. The true kernel is an algebraic function; and though the exponential approximation may be higher than the algebraic function in some places, the exponential must inevitably fall below the true values at sufficiently large values of $(s-t)$. This fact does not cause much trouble if $s<4$ and if there is no reflection from the bottom (as can be seen by the good agreement in Figs. 4 and 6) because the portion of the kernel near $(s-t)=0$ is always in use and exercises a predominant

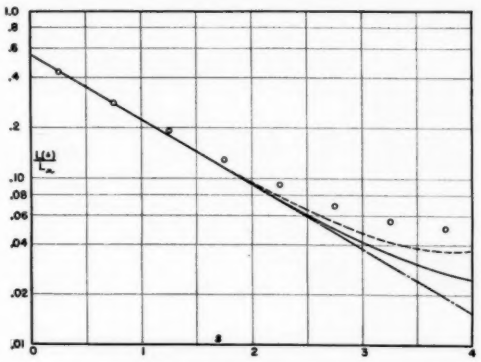


Fig. 7. Lightwell with reflecting bottom. ($w=1$, $l=4$, $\rho_1=\rho_2=0.80$). ○ Manning and White; —— Eq. (15); —·— Eq. (10); – – – Eq. (20).

effect. But when a highly reflecting bottom is introduced at $l=4$, the initial illumination on the bottom, as calculated by the exponential formula, is much too low. Thus, the effect of the bottom according to the exponential relation is smaller than the actual effect.

6. Another Approximation

In the previous work, the direct illumination on the bottom of the lightwell has been obtained by use of an approximate (exponential) formula. It is now proposed to use the exact expression for the direct illumination of the bottom, but to keep the exponential kernel as before.

The integral equation is[7]

$$L_1(s)=L_0(s)+\rho_1\int_0^l K(s,t)L_1(t)dt. \quad (17)$$

For a lightwell with open top and reflecting bottom, equations given in the previous paper[7] reduce to

$$L_0(s)=\frac{\rho_1L_u}{2}[e^{-as}+\rho_3K_{23}e^{-a(l-s)}], \quad (18)$$

$$K(s,t)=\frac{a}{2}[e^{-a|s-t|}+\rho_3e^{-2al}e^{a(s+t)}], \quad (19)$$

where

K_{23} = average illumination of the bottom, per unit luminosity of sky.

The solution is of the form

$$L_1(s)=B\cosh ks+C\sinh ks. \quad (20)$$

Substitution of Eq. (20) in Eq. (17) allows the evaluation of the coefficients, which are found to be

$$B=\rho_1 L_u\left\{\frac{(1-\rho_1)^{\frac{1}{2}}\rho_3[K_{23}-e^{-al}]+(1-\rho_1)^{\frac{1}{2}}(1+\rho_3)\cosh kl+(1-\rho_3)\sinh kl}{2(1-\rho_1)^{\frac{1}{2}}\cosh kl+[(1-\rho_3)+(1-\rho_1)(1+\rho_3)]\sinh kl}\right\}, \tag{21}$$

$$C=-\rho_1 L_u\left\{\frac{-\rho_3[K_{23}-e^{-al}]+(1-\rho_1)^{\frac{1}{2}}(1+\rho_3)\sinh kl+(1-\rho_3)\cosh kl}{2(1-\rho_1)^{\frac{1}{2}}\cosh kl+[(1-\rho_3)+(1-\rho_1)(1+\rho_3)]\sinh kl}\right\}. \tag{22}$$

If K_{23} has its previous value of e^{-al}, B and C reduce to the expression given in Eq. (16). The exact formula is

$$K_{23}=\frac{1}{\pi}\left\{l^2\ln\left[\frac{(1+l^2)^2}{l^2(2+l^2)}\right]-4l\tan^{-1}(1/l)+4(l^2+1)^{\frac{1}{2}}\tan^{-1}(l^2+1)^{-\frac{1}{2}}\right\}. \tag{23}$$

If l is large, as in the $1\times1\times4$ lightwell, the two inverse tangents cancel and Eq. (23) becomes

$$K_{23}\cong\frac{l^2}{\pi}\ln(l^{-2})\cong\frac{l^2}{\pi}\left[\frac{1}{l^4}-\frac{1}{2}\left(\frac{1}{l^4}\right)^2+\cdots\right]\cong\frac{1}{\pi l^2}, \tag{24}$$

which is the ordinary inverse-square relation. Also, if l is large a form of Eq. (20) more convenient for use in computation is

$$\frac{L_1(s)}{L_u}=\frac{\rho_1}{2W}\{[\rho[K_{23}-e^{-al}][1+(1-\rho_1)^{\frac{1}{2}}]+[(1-\rho_1)^{\frac{1}{2}}(1+\rho_3)-(1-\rho_3)]e^{-kl}]e^{ks}$$
$$+[\rho[K_{23}-e^{-al}][-1+(1-\rho_1)^{\frac{1}{2}}]+[(1-\rho_1)^{\frac{1}{2}}(1+\rho_3)+(1-\rho_3)]e^{kl}]e^{-ks}\}, \tag{25}$$

where

$$W=2(1-\rho_1)^{\frac{1}{2}}\cosh kl+[(1-\rho_3)+(1-\rho_1)(1+\rho_3)]\sinh kl\cong\{(1-\rho_1)^{\frac{1}{2}}+\tfrac{1}{2}[(1-\rho_3)+(1-\rho_1)(1+\rho_3)]\}e^{kl}.$$

Equation (25) was used in computing the dotted curve of Fig. 7. The luminosity of the lower part of the wall is raised, but not enough to agree with the values of Manning and White. Thus it appears that the use of Eqs. (21), (22), and (23) is hardly advisable. The previous method of Eqs. (15) and (16) is a self-consistent one, since it uses the exponential approximation throughout, and the formulas are simpler.

The discrepancy of Fig. 7 is not important in practice. Though the maximum discrepancy is 46 percent, based on the Manning-White value at $s=3.75$, the discrepancy is only 4.2 percent of the wall luminosity at the top of the lightwell and only 2.3 percent of the sky luminosity. The illumination in the lower rooms facing the lightwell is produced largely by direct light from the sky, with some help from the upper walls of the court. The lower part of the lightwell does little more than help to illuminate the rear wall of the room, and this illumination is so low in any case that a slight additional pessimism in the calculated results is of no consequence. Thus it appears that Eqs. (15) and (16) are satisfactory in all ordinary cases.

7. Allowable Depth of Lightwell

It is hoped that the results of this paper will help the architect in deciding on the dimensions of lightwells and the spacing between buildings and will help the city planner in deciding on desirable street widths and building heights. The first step is to establish a criterion for the minimum allowable ratio of wall luminosity to sky luminosity, $L_1(l)/L_u$. Obviously, this criterion will depend on a great number of factors, including the dimensions of the rooms to be lighted, the window sizes, and the

reflection factor of the room walls.[9] A study of these variables is beyond the scope of this paper, so the assumption will be made that a reasonable criterion has been established:

$$L_1(l)/L_u = c = \text{const.}$$

For a given value of c, what is the maximum allowable depth of a lightwell? The question can be answered by using Eqs. (5), (6), and (7). For $s=l$,

$$\frac{c}{\rho_1}[2(1-\rho_1)^{\frac{1}{2}}\cosh kl+(2-\rho_1)\sinh kl]=[(1-\rho_1)^{\frac{1}{2}}\cosh kl+\sinh kl]\cosh kl - [\cosh kl+(1-\rho_1)^{\frac{1}{2}}\sinh kl]\sinh kl$$

or

$$\cosh kl+\frac{2-\rho_1}{2(1-\rho_1)^{\frac{1}{2}}}\sinh kl=\frac{\rho_1}{2c}. \tag{26}$$

Given the values of ρ_1 and c, a few trials with the aid of a table of hyperbolic functions will locate the value of kl. Then the maximum allowable depth of lightwell is

$$l_{\max}=(kl)/a(1-\rho_1)^{\frac{1}{2}}. \tag{27}$$

The lower curve of Fig. 8 was calculated by use of Eq. (26) with $c=0.10$.

The foregoing treatment applies to lightwells with nonreflecting bottoms. Similar results are obtained when reflection from the bottom is taken into account. Equation (15) gives

$$\cosh kl+\frac{(1-\rho_3)+(1-\rho_1)(1+\rho_3)}{2(1-\rho_1)^{\frac{1}{2}}}\sinh kl = \frac{\rho_1(1+\rho_3)}{2c}. \tag{28}$$

In the special case where the reflection factor of the bottom is equal to the average reflection factor of the walls,

$$\cosh kl+\frac{(1-\rho_1)(2+\rho_1)}{2(1-\rho_1)^{\frac{1}{2}}}\sinh kl=\frac{\rho_1(1+\rho_1)}{2c}. \tag{28a}$$

The upper curve of Fig. 8 was obtained by use of Eq. (28a) with $c=0.10$. The improvement obtained by using high reflection factors is rather striking. But high reflection factors are not ordinarily obtainable because of the windows. Suppose that a glazed white tile is used having a reflection factor of 0.80. Meacock and Lambert have shown that the results are very nearly the same as with a matt surface having the same reflection factor. Suppose that the windows are 2.1 m square and are separated, both horizontally and vertically, by 1.5 m of tiled wall. Then the average reflection factor is

$$\rho_1=\frac{8.55(0.80)+4.41(0.10)}{12.96}=0.562,$$

where the reflection factor of the glass is taken as 0.10. Reference to Fig. 8 shows that, even with these highly reflecting tiles, the depth of a square lightwell ($a=2$) should not exceed approximately its width. For a 1×2 section ($a=1.5$), the depth can be increased 33 percent; while for the space between long parallel buildings ($a=1.0$), l can be 100 percent greater than for a square lightwell having the same spacing between opposite walls.

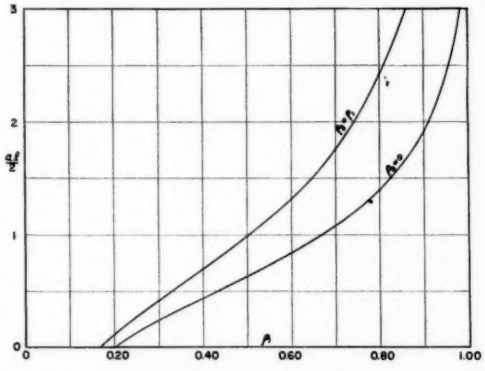


FIG. 8. Maximum depths of a lightwell to give $L(s)/L_u = 0.10$ on the walls near the bottom.

[9] K. Y. Tang, "A method of predicting illumination from light courts," Ohio State Univ. Eng. Exp. Station, Bull. No. 47 (1928).

[10] H. H. Higbie and A. D. Moore, "Prediction of natural illumination in interiors and on walls of buildings," C.I.E. Proc., p. 1149 (1931).

8. Summary

The solution of an integral equation is employed in expressing the luminosity distribution in lightwells. Reasonably simple expressions are given for lightwells having any dimensions and any reflection factors for sides and bottom. The values calculated from these equations are compared with experimental results and with values previously obtained by a laborious process of numerical approximation. The new formulas are found to give results that are sufficiently precise for ordinary engineering purposes. Thus the equations constitute a new tool for the engineer and the architect, helping them to obtain a logical daylighting design without resorting to experimental work.

APRIL, 1941 J. O. S. A. VOLUME 31

Calculation of the Light Distribution in Lightwells

WALTER T. WHITE
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received December 19, 1940)

1. Introduction

THE problem of determining the luminosity of the walls of a lightwell which is illuminated by a uniform sky has been considered by a number of investigators, using both analytical and experimental methods.[1–6] The problem can be formulated analytically as an integral equation, but in practically all cases a solution of the integral equation can be obtained only by approximate or by numerical methods.

In the present paper the luminosity of the walls in finite lightwells illuminated by a sky of uniform luminosity will be calculated by numerical solutions of the integral equations involved. The following lightwells are considered: (a) Right circular cylinder of unit diameter, 0.5 unit depth, and unit reflection factor. (b) Rectangular parallelepiped with cross sections 1×1, 1×2, and 1×5 units; depth equal to 4 units; and reflection factors equal to 1.0, 0.8, 0.5, and 0.25. All of the lightwells will be assumed to have a perfectly absorbing bottom. The numerical solutions thus obtained are general in that they hold for all lightwells whose dimensional ratios are equal to the ones considered, irrespective of the actual size.

2. Formulation of the Integral Equation

In the interreflection problem the following assumptions are made: (a) All reflecting surfaces are perfectly diffusing. (b) The luminosity at any point can be expressed as a function of a single variable. (c) The radiation is homogeneous. (d) The medium between the reflecting surfaces is transparent and nondissipative. (e) The sky is of uniform luminosity. With these assumptions the actual luminosity $L(s)$ in a lightwell with walls having a uniform reflection factor ρ and a perfectly absorbing bottom is[6]

$$L(s)=L_0(s)+\rho\int_0^s K(s-t)L(t)dt+\rho\int_s^l K(t-s)L(t)dt, \quad (1)$$

where $L_0(s)$ = initial luminosity due to the source of illumination; $K(s,t)$ = the kernel of the integral equation, or the illumination per unit distance at s due to a source of unit luminosity

[1] H. Buckley, "On the radiation from the inside of a circular cylinder," Phil. Mag. **4**, 753 (1927).
[2] A. D. Moore, "Inter-reflection by the increment method as applied to a light court," Trans. I. E. S. **24**, 629 (1929).
[3] E. W. Manning and A. D. White, "Brightness distribution due to interreflections in a light court," Trans. I. E. S. **25**, 663 (1930).
[4] H. F. Meacock and G. E. V. Lambert, "The efficiency of lightwells," Dept. Sci. Ind. Research, London, Illumin. Research, Tech. Pap. No. 11 (1930). Proc. C. I. E. (1931), p. 1236.
[5] H. C. Hottel and J. D. Keller, "Effect of reradiation on heat transmission in furnaces and through openings," A. S. M. E., presented at Annual Meeting, New York (1932).
[6] P. Moon, "On interreflections," J. Opt. Soc. Am. **30**, 195 (1940).

along unit distance at t; s = depth in lightwell at which luminosity is considered; t = parameter of integration, and l = total depth of lightwell.

This relation expresses the fact that the actual luminosity at the depth s is equal to the initial luminosity plus the luminosity due to interreflections from the side walls of the lightwell above and below the section considered. The expression is an integral equation of the second kind, Fredholm type, and is to be solved for the unknown luminosity function $L(s)$.

In most cases an exact solution of the integral equation by analysis is impossible, and, therefore, either approximate or numerical methods must be used. As generally applied these methods are very tedious and time-consuming. Recourse to some sort of special calculating machine to reduce the burden of the numerical work would make the solution of the integral equation practicable, and such a calculating machine is available for evaluating integral equations of the form of Eq. (1).

3. A Machine for Evaluating a Parametric Product Integral

Recently a machine which facilitates the numerical evaluation of integral equations was developed at the Massachusetts Institute of Technology. The *cinema integraph*,[7] as it is called, can be used to obtain a numerical solution of a parametric product integral of the general form

$$F(t)=\int_a^{b \text{ or } t} f(x, t)g(x)dx. \tag{2}$$

The machine involves the use of radiation from a line source of light, which radiation is transmitted through transparent areas on two separate photographic films. The transparent areas on the two films are of the form of the functions $f(x, t)$ and $g(x)$ in Eq. (2), and the total radiation transmitted through the two films is proportional to the integral being evaluated. This integrating radiation is collected by an optical system and impressed upon a photoelectric cell. The response of the cell is proportional to the numerical value of the integral, and convenient methods are available for calibrating the machine and placing the answer in the desired system of units. The answer can be read from a dial on the machine; or in case one of the functions is of the form $f(t-x)$, the machine plots on standard graph paper the result $F(t)$ as a continuous function of the parameter t. The integral can be evaluated to an accuracy of about one percent.

[7] H. L. Hazen and G. S. Brown, "The cinema integraph, a machine for evaluating a parametric product integral," J. Frank. Inst. **230**, 19 (1940).

4. Evaluation of an Integral Equation with the Cinema Integraph

A method for using the cinema integraph to solve problems on interreflections follows from a consideration of the general integral equation of the second kind, Fredholm type,

$$u(t)=f(t)+\lambda\int_a^b K(x, t)u(x)dx, \tag{3}$$

where $u(t)$ is unknown, $f(t)$ = known exciting function, $K(x, t)$ = known kernel, λ = known constant, and a and b are known limits.

The solution can be obtained by computing the series of functions defined by the equations:[8,9]

$$\left.\begin{aligned} u_1(t)&=f(t)+\lambda\int_a^b K(x, t)u_0(x)dx\\ u_2(t)&=f(t)+\lambda\int_a^b K(x, t)u_1(x)dx\\ &\cdots\\ u_n(t)&=f(t)+\lambda\int_a^b K(x, t)u_{n-1}(x)dx \end{aligned}\right\}, \tag{4}$$

where $u_0(x)$ is arbitrarily chosen. Regardless of the initial choice of $u_0(t)$, it can be shown that

$$\lim_{n\to\infty} u_n(t)=u(t). \tag{5}$$

The value of the method lies in the fact that if a reasonably good approximation is used as the first choice for $u_0(t)$, the convergence is very rapid and $u(t)$ frequently can be obtained in no more than two to seven trials. When the cinema integraph is used to evaluate the integrals, this

[8] M. Bôcher, *An Introduction to the Study of Integral Equations* (Cambridge, University Press, 1926), second edition.

[9] W. V. Lovitt, *Linear Integral Equations* (McGraw-Hill, New York, 1924).

method of successive approximations is entirely practicable and does not require a prohibitive length of time. The process is continued until the functions $u_n(t)$ and $u_{n-1}(t)$ in Eqs. (4) are found to be identical, within the limits of error of the machine. Other methods, such as the use of the Neumann series,[9,10] or the "reciprocal kernel,"[10,11] can be used to advantage with the cinema integraph for solving the integral equation.

5. Perturbation Method for High Accuracy

In order to obtain high accuracy, a perturbation method can be used with success. An approximate solution of the integral equation is determined with the cinema integraph, from which an analytic function (usually consisting of straight line segments) can be formed to be used in evaluating the greater part of the integral by ordinary analysis. The difference between the exact solution and the analytic function represents a correction function which is the unknown in a new integral equation. Since the correction function represents a small part of the exact solution, a one percent error in its evaluation will hardly affect the final result.

When the reflection factor is unity, the luminosity function of Eq. (1) can be expressed as

$$L(s)/L_u = ms+b+H(s), \tag{6}$$

where L_u = luminosity of uniform sky, $(ms+b)$ = known linear variation, and $H(s)$ = unknown correction function. Substitution of Eq. (6) in Eq. (1) gives

$$\begin{aligned} ms+b+H(s) &= L_0(s)/L_u \\ &+\int_0^s K(s-t)H(t)dt+\int_s^l K(t-s)H(t)dt \\ &+m\int_0^s tK(s-t)dt+m\int_s^l tK(t-s)dt \\ &+b\int_0^s K(s-t)dt+b\int_s^l K(t-s)dt. \end{aligned} \tag{7}$$

From the literature,[5] the following relations are known:

$$K(s) = -L_0'(s)/\rho L_u, \tag{8}$$

$$\frac{L_0(s)}{\rho L_u} = -\frac{w}{2(w+1)}\cdot F'(s), \tag{9}$$

where L_u is the luminosity of the uniform sky, the factor w is the ratio of the long to the short side of the cross section of the lightwell (equals unity for a circle), and the function $F(s)$ is an *area* factor dependent on the solid angle filled by the luminous flux from an "equivalent sky" to the walls of the lightwell. $F(s)$ is equal to the fraction of the luminous flux from the equivalent sky which is intercepted within the bounds of the lightwell. The "equivalent sky" may be thought of as a flat diaphragm of unit luminosity placed at the top of the lightwell. Thus at the top of the lightwell, where $s=0$, all the luminous flux from the equivalent sky is intercepted within the bounds of the lightwell and $F(0)$ is unity. Derivations of the functions $K(s)$, $L_0(s)$, and $F(s)$ have been reported in the literature[5,6] and are listed in Table I for lightwells of both circular and rectangular cross sections.

Substitution of Eqs. (8) and (9) into Eq. (7) and application of the fundamental theorem on integrals

$$\int udv = uv - \int vdu \tag{10}$$

gives

$$\begin{aligned} H(s) = G(s) &+ \int_0^s K(s-t)H(t)dt \\ &+ \int_s^l K(t-s)H(t)dt. \end{aligned} \tag{11}$$

The function $G(s)$ is defined as

$$\begin{aligned} G(s) = &\frac{mw}{2(w+1)}[F(s)-F(l-s)] \\ &+(1-b)\frac{L_0(s)}{L_u}-(ml+b)\frac{L_0(l-s)}{L_u}. \end{aligned} \tag{12}$$

Equation (11) can be solved on the cinema integraph for $H(s)$ and the summation of $H(s)$ and the known linear function of Eq. (6) gives the desired luminosity function $L(s)/L_u$.

[10] W. R. Hedeman, Jr., *The Numerical Solution of Integral Equations on the Cinema Integraph*, Sc. D. Thesis, Massachusetts Institute of Technology (1939).

[11] V. Volterra, *Leçons sur les Équations Intégrales* (Gauthier-Villars, Paris, 1913); Rice Inst. Pamphlet, Vol. 7 (1920).

TABLE I.

FORM	$F(s)$	$\frac{L_0(s)}{\rho L_u}=-\frac{w}{2(w+1)}\cdot F'(s)$	$K(s)=-\frac{L_0'(s)}{\rho L_u}$
Circular cylinder $(w=1)$	$1+2s^2-2s(s^2+1)^{\frac{1}{2}}$	$\frac{2s^2+1}{2(s^2+1)^{\frac{1}{2}}}-s$	$1-\frac{s(2s^2+3)}{s(s^2+1)^{\frac{1}{2}}}$
Rectangular parallelepiped* $(1\times w)$	$\frac{2}{\pi}\Big\{\frac{(s^2+w^2)^{\frac{1}{2}}}{w}\tan^{-1}\frac{1}{(s^2+w^2)^{\frac{1}{2}}}+(s^2+1)^{\frac{1}{2}}\tan^{-1}\frac{w}{(s^2+1)^{\frac{1}{2}}}-\frac{s^2}{w}\Big[\frac{1}{s}\tan^{-1}\frac{1}{s}+\frac{w}{s}\tan^{-1}\frac{w}{s}+\frac{1}{2}\ln\frac{s^2(s^2+w^2+1)}{(s^2+1)(s^2+w^2)}\Big]\Big\}$	$\frac{s}{\pi(w+1)}\Big[\ln\frac{s^2(s^2+w^2+1)}{(s^2+1)(s^2+w^2)}+\frac{1}{s}\tan^{-1}\frac{1}{s}+\frac{w}{s}\tan^{-1}\frac{w}{s}-\frac{1}{(s^2+w^2)^{\frac{1}{2}}}\tan^{-1}\frac{1}{(s^2+w^2)^{\frac{1}{2}}}-\frac{w}{(s^2+1)^{\frac{1}{2}}}\tan^{-1}\frac{w}{(s^2+1)^{\frac{1}{2}}}\Big]$	$\frac{1}{\pi(w+1)}\Big[\ln\frac{(s^2+1)(s^2+w^2)}{s^2(s^2+w^2+1)}+\frac{w^2}{(s^2+w^2)^{\frac{1}{2}}}\tan^{-1}\frac{1}{(s^2+w^2)^{\frac{1}{2}}}+\frac{w}{(s^2+1)^{\frac{1}{2}}}\tan^{-1}\frac{w}{(s^2+1)^{\frac{1}{2}}}\Big]$

* Approximate, since the value of the function at s is taken as the arithmetic mean of the actual values around the cross section of the rectangle.

6. PROPERTIES OF LUMINOSITY FUNCTION WHEN REFLECTION FACTOR IS UNITY

The luminosity function for unit reflection factor has a value at the mid-section of the lightwell equal to one-half the luminosity of the sky, and is a skew-symmetrical function about that section. The proof of these properties follows from the demonstration that such a function satisfies the integral equation. Let

$$L(s)=L_u/2+P(s), \tag{13}$$

where $P(s)$ is defined as being odd about the mid-section, that is,

$$P(l/2)=0, \tag{14}$$

and

$$P(s)=-P(l-s). \tag{15}$$

It is required to demonstrate that the substitution of Eq. (13) in Eq. (1) gives

$$L(s)-L_u/2=-[L(l-s)-L_u/2], \tag{16}$$

which is the condition for skew-symmetry about the mid-section. Also, substitution of $s=l/2$ gives

$$L(l/2)=L_u/2. \tag{16a}$$

Substitution of Eq. (13) in Eq. (1) for sections s and $(l-s)$ gives

$$L(s)=L_0(s)+(L_u/2)\int_0^s K(s-t)dt+(L_u/2)\int_s^l K(t-s)dt+\int_0^l K|s-t|P(t)dt, \tag{17}$$

and

$$L(l-s)=L_0(l-s)+(L_u/2)\int_0^{l-s} K(l-s-t)dt+(L_u/2)\int_{l-s}^l K(t-l+s)dt+\int_0^l K|l-s-t|P(t)dt. \tag{18}$$

Substitution of Eq. (8) ($\rho=$ unity) into the first two integrals of Eqs. (17) and (18) gives

$$L(s)=L_0(0)+\tfrac{1}{2}[L_0(s)-L_0(l-s)]+\int_0^l K|s-t|P(t)dt, \tag{17a}$$

and

$$L(l-s)=L_0(0)-\tfrac{1}{2}[L_0(s)-L_0(l-s)]+\int_0^l K|l-s-t|P(t)dt. \tag{18a}$$

Since the kernel is symmetric,[6,11] that is,

$$K(s-t)=K(t-s),$$

and since by hypothesis $P(s)$ is odd about $s=l/2$, it follows

$$\int_0^l K|s-t|P(t)dt=-\int_0^l K|l-s-t|P(t)dt. \tag{19}$$

Also, for unit reflection factor, the initial luminosity at the top of the lightwell is equal to

TABLE II.

DEPTH s	$H(s)$	$L(s)/L_u$	HOTTEL AND KELLER $L(s)/L_u$
0	8.14×10^{-4}	0.673814	
$\frac{1}{32}$	13.02	0.652552	0.65239
$\frac{1}{16}$	15.99	0.631099	
$\frac{3}{32}$	17.29*	0.609479	0.60937
$\frac{1}{8}$	17.44	0.587744	
$\frac{5}{32}$	16.62*	0.565912	0.56583
$\frac{3}{16}$	14.72	0.543972	
$\frac{7}{32}$	12.45*	0.521995	0.52204
$\frac{1}{4}$	9.99	0.499999	
$\frac{9}{32}$	7.60*	0.478010	0.47796
$\frac{5}{16}$	5.25	0.456025	
$\frac{11}{32}$	3.46*	0.434086	0.43417
$\frac{3}{8}$	2.56	0.412256	
$\frac{13}{32}$	2.75*	0.390515	0.39063
$\frac{7}{16}$	3.99	0.368899	
$\frac{15}{32}$	6.96	0.347446	0.34761
$\frac{1}{2}$	11.88×10^{-4}	0.326188	

* Determined from graphical plot of $H(s)$.

one-half the luminosity of the sky, that is,

$$L_0(0)=L_u/2. \tag{20}$$

Substitution of Eqs. (19) and (20) in Eqs. (17a) and (18a) and the addition of these two equations gives

$$L(s)-L_u/2=-[L(l-s)-L_u/2]; \tag{16}$$

and at the section $s=l/2$,

$$L(l/2)=L_u/2. \tag{16a}$$

Therefore, a luminosity function with properties as defined by Eq. (13) satisfies the integral equation.

These properties of the luminosity function for unity reflection factor are of considerable value in checking the accuracy of any solution of the integral equation obtained by either approximate or numerical methods. When a method is developed which gives a solution satisfying these properties, further uses of the method will require this particular integral equation being solved only for the upper or lower half of the lightwell.

7. LUMINOSITY IN A FINITE RIGHT CIRCULAR CYLINDER

In a problem of heat re-radiation, Hottel and Keller[5] determined an emissivity function analogous to luminosity for the case of a finite cylinder receiving radiation from a uniform source. The reflection factor of the walls was unity and the bottom was a perfectly absorbing body. The problem was solved by replacing the integral equation with a set of linear simultaneous equations, and the result showed that the luminosity varied *almost linearly* from the top to the bottom of the cylinder. It was the purpose of this study to determine precisely how the luminosity function differed from linearity.

In order to obtain high accuracy, the perturbation method of paragraph 5 was employed. Based on the results of Hottel and Keller for a cylinder of depth $l=0.5$, the constants for the linear part of Eq. (6) were taken as

$$m=-0.696, \quad \text{and} \quad b=0.673.$$

The function $G(s)$ was calculated by Eq. (12),

TABLE III. *Luminosity of the walls in rectangular parallelepiped lightwells.*

DEPTH s	CROSS SECTION 1×1, $w=1$				CROSS SECTION 1×2, $w=2$				CROSS SECTION 1×5, $w=5$			
	$\frac{L(s)}{L_u}$ $\rho=1.0$	$\frac{L(s)}{L_u}$ $\rho=0.8$	$\frac{L(s)}{L_u}$ $\rho=0.5$	$\frac{L(s)}{L_u}$ $\rho=0.25$	$\frac{L(s)}{L_u}$ $\rho=1.0$	$\frac{L(s)}{L_u}$ $\rho=0.8$	$\frac{L(s)}{L_u}$ $\rho=0.5$	$\frac{L(s)}{L_u}$ $\rho=0.25$	$\frac{L(s)}{L_u}$ $\rho=1.0$	$\frac{L(s)}{L_u}$ $\rho=0.8$	$\frac{L(s)}{L_u}$ $\rho=0.5$	$\frac{L(s)}{L_u}$ $\rho=0.25$
0	0.8972	0.550	0.292	0.134	0.8732	0.547	0.292	0.134	0.8549	0.549	0.292	0.134
0.25	0.8426	0.427	0.192	0.080	0.8200	0.445	0.211	0.090	0.8065	0.466	0.226	0.098
0.50	0.7929	0.343	0.134	0.051	0.7719	0.372	0.158	0.064	0.7598	0.398	0.177	0.072
0.75	0.7441	0.274	0.094	0.034	0.7248	0.314	0.121	0.046	0.7145	0.345	0.139	0.054
1.00	0.6954	0.221	0.068	0.023	0.6787	0.266	0.094	0.033	0.6702	0.298	0.111	0.040
1.25	0.6464	0.178	0.050	0.016	0.6337	0.226	0.073	0.024	0.6268	0.259	0.090	0.032
1.50	0.5978	0.144	0.036	0.011	0.5888	0.193	0.057	0.018	0.5841	0.226	0.073	0.024
1.75	0.5489	0.116			0.5442				0.5419			
2.00	0.4999	0.094	0.021	0.006	0.4998	0.142	0.037	0.011	0.4999	0.174	0.050	0.016
2.25	0.4515	0.077			0.4555				0.4579			
2.50	0.4027	0.062	0.013	0.003	0.4110	0.105	0.024	0.007	0.4157	0.134	0.035	0.011
2.75	0.3539	0.050			0.3663				0.3731			
3.00	0.3047	0.040	0.008	0.002	0.3211	0.076	0.017	0.005	0.3297	0.101	0.025	0.008
3.25	0.2560	0.031			0.2749				0.2854			
3.50	0.2068	0.025	0.005	0.001	0.2278	0.051	0.010	0.004	0.2403	0.071	0.018	0.006
3.75	0.1568	0.018			0.1798				0.1937			
4.00	0.1026	0.013	0.003	0.001	0.1266	0.029	0.007	0.002	0.1451	0.044	0.012	0.004

using the functions in Table I for a circular cylinder. Equation (11) was evaluated on the cinema integraph for the correction function $H(s)$, and the luminosity function was determined from Eq. (6). These results and the results of Hottel and Keller are given in Table II. The values of $L(s)/L_u$ from this solution are estimated to have an accuracy better than $\pm 1\times 10^{-5}$.

An examination of the luminosity function in Table II shows two interesting properties. First, the function is equal to exactly one-half the luminosity of the sky at the section $s=l/2$; and secondly, it is skew-symmetrical about this section (the slight deviations from an exact agreement with these statements are due to errors in the computational methods). The observation of these properties suggested the proof in paragraph 6.

8. Luminosity in Finite Rectangular Parallelepipeds

The rectangular parallelepiped lightwells with unit reflection factor were considered by the perturbation method. An approximate solution of Eq. (1) was obtained by the cinema integraph for the lightwell with square cross section. This solution determined constants for the linear

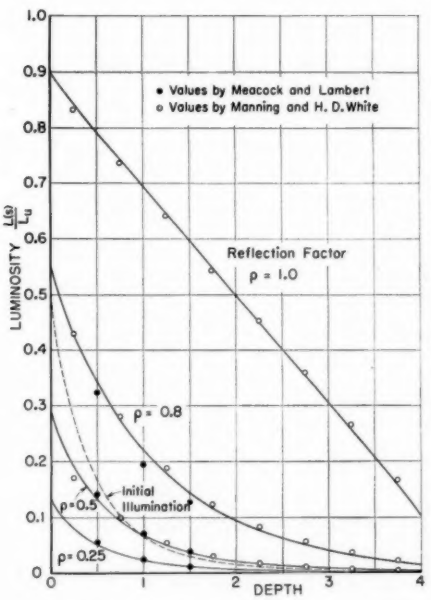


Fig. 1. Luminosity of the walls in a rectangular parallelepiped lightwell, cross section 1×1.

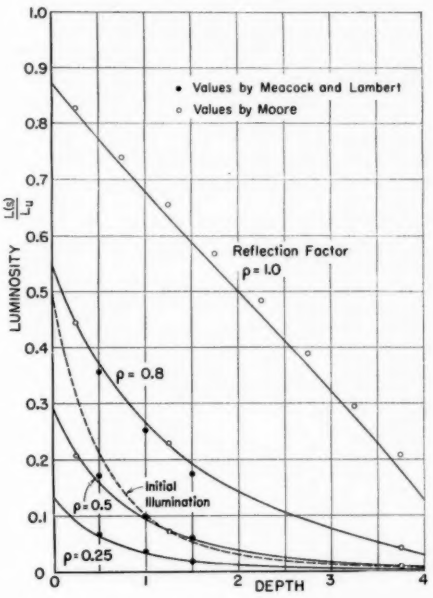


Fig. 2. Luminosity of the walls in a rectangular parallelepiped lightwell, cross section 1×2.

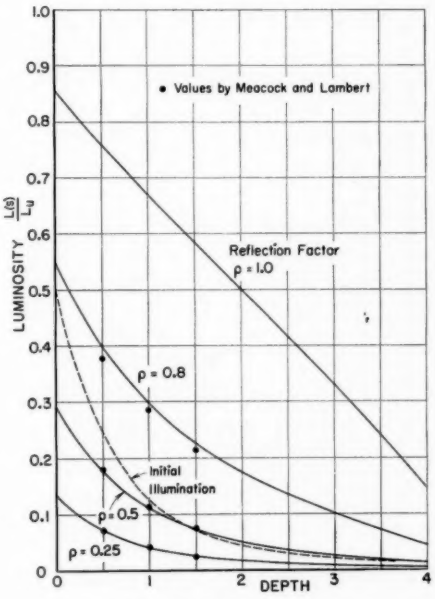


Fig. 3. Luminosity of the walls in a rectangular parallelepiped lightwell, cross section 1×5.

variation, and these constants were used with each of the three rectangular cross sections. The constants are

$$m=-0.184, \quad \text{and} \quad b=0.84.$$

The function $G(s)$ was calculated by Eq. (12) and Table I for cross sections with w equal, respectively, to 1, 2 and 5. Equation (11) was evaluated by the cinema integraph in each case for $H(s)$, and the summation of each correction function with the linear variation gave the desired luminosity function. The final results are given for the three cases in Table III and are plotted in Figs. 1 to 3. It is estimated that the values for $L(s)/L_u$ are accurate to better than $\pm 1\times 10^{-3}$.

The accuracy is not as high as for the circular cylinder, since the assumption of the same linear variation for the three cases left a greater part of the integral to be evaluated by the cinema integraph. Nevertheless, the fact that the luminosity function has a value of one-half the luminosity of the sky at the mid-section and its skew-symmetry about that point is definitely shown.

The actual luminosity in a rectangular lightwell varies slightly around a cross section. Since the results given in Table III represent *average values* around the cross section, they actually are less accurate than results for the circular case where no averaging is necessary.

The rectangular lightwells with reflection factors of 0.8, 0.5, and 0.25 were studied by solving Eq. (1) directly with the cinema integraph. These results are given in Table III and are plotted in Figs. 1 to 3. It is estimated that the results in each case are accurate to within one percent of the maximum value of the curve.

9. Comparison with Results of Other Investigators

Other investigators have considered the rectangular lightwells and presented results of their studies.[2–4] Some of these results are plotted in Figs. 1 to 3.

Meacock and Lambert studied a variety of lightwells by experimental methods using models. Their results were the most complete for a lightwell ten units deep, and it was from these values that the points plotted in the figures were selected. In the range of comparison, their results show that one would obtain values only very slightly lower for a lightwell four units deep. Therefore, the comparison is significant. Good agreement is observed for reflection factors 0.25 and 0.5, but for 0.8 reflection factor the Meacock and Lambert values are low for all three cross sections.

Moore, Manning and H. D. White have calculated the luminosity in the square and 1×2 lightwells by an "increment method."[2,3] Some of their values are plotted on Figs. 1 and 2, and good agreement is observed except for the 1×2 lightwell with unit reflection factor. In this case the results of Moore are definitely high, since they do not pass through the value $L_u/2$ at the mid-section.

10. Solution of Integral Equations by Simultaneous Equation Method

The perturbation method of dividing the unknown function of the integral equation into a linear part and a correction function, as by Eq. (6), suggests an interesting method of solving integral equations by a simultaneous equation method. Briefly, the method is to assume the solution of the integral equation to be a function made up of straight line segments whose slopes and intercepts are unknown. The substitution of this function into the integral equation leads to a set of linear simultaneous equations which can be solved for the unknown slopes and intercepts.

An application of the method to the solution of the integral equation for the lightwell of square cross section and unit reflection factor will serve as an illustrative example. Let the unknown luminosity function be given by

$$L(s)/L_u=Q(s)+H(s), \tag{21}$$

where

$$\left.\begin{aligned} Q(s)&=m_1 s+b_1 \quad \text{on} \quad 0\leqslant s\leqslant l/2\\ &=m_2 s+b_2 \quad \text{on} \quad l/2\leqslant s\leqslant l \end{aligned}\right\}, \tag{22}$$

represent the two assumed straight line segments. $H(s)$ is the correction function between the exact $L(s)/L_u$ and the straight line segments. Substitution of Eq. (22) into Eq. (1) gives a new integral equation with $H(s)$ as the unknown.

For the interval

$$0 \leq s \leq l/2,$$

$$H(s) = \frac{L_0(s)}{L_u} + m_1\left\{\frac{w}{2(w+1)}[F(s) - F(l/2-s)] - \frac{l}{2L_u}L_0(l/2-s)\right\} - \frac{b_1}{L_u}[L_0(s)+L_0(l/2-s)] + m_2\left\{\frac{w}{2(w+1)}[F(l/2-s) - F(l-s)] + \frac{l}{L_u}\frac{[L_0(l/2-s)}{2} - L_0(l-s)]\right\} + \frac{b_2}{L_u}[L_0(l/2-s) - L_0(l-s)] + \int_0^l K|t-s|H(t)dt. \quad (23)$$

Now, if the straight line segments are a reasonably good approximation of the unknown function,

$$H(s) - \int_0^l K|t-s|H(t)dt \leqq 0, \quad (24)$$

which may be neglected. (This would be exactly true if an infinitely large number of straight line segments were used.) The approximation of Eq. (24) leads to a linear equation with four unknowns m_1, b_1, m_2, and b_2.

A relation similar to Eq. (23) can be formed for the interval

$$l/2 \leqslant s \leqslant l.$$

The constant terms and coefficients of the unknowns can be determined by reference to the functions in Table I for four values of the parameter s. In the case actually solved, values of s equal to zero, $l/2$, and l were used. A fourth equation was obtained by applying the restrictive condition that the two linear segments have the same value at $l/2$.

The four equations thus formed are:

$$\left.\begin{aligned} 0.20955m_1 - 0.51490b_1 + 0.02859m_2 + 0.01261b_2 &= -0.50000 \\ -1.23935m_1 - 0.51490b_1 + 1.17975m_2 + 0.48510b_2 &= -0.01490 \\ 0.02184m_1 + 0.01261b_1 - 2.26914m_2 - 0.51490b_2 &= -0.00229 \\ 2m_1 + b_1 - 2m_2 - b_2 &= 0 \end{aligned}\right\}. \quad (25)$$

The simultaneous solution of this set of equations gives

$$m_1 = -0.20026 \quad m_2 = -0.20025$$
$$b_1 = 0.90051 \quad b_2 = 0.90048.$$

Luminosity values from the solution are tabulated below (Table IV) and compared with the more exact values from Table III.

It is evident that the simultaneous equation method gives good results in the example considered. And further, a single straight line approximation will give an equally satisfactory result, since the slopes and intercepts are almost identical.

For a single straight line approximation with coefficients and constant terms evaluated at

$$s = 0, \quad \text{and} \quad s = l,$$

the unknowns m and b can be calculated readily. The results are

$$m = \frac{-1}{[wL_u/(w+1)]\cdot[(F(0)-F(l))/(L_0(0)-L_0(l))]+l}, \quad (26)$$

and

$$b = \tfrac{1}{2}\left[1 + \frac{1}{[wL_u/l(w+1)]\cdot[(F(0)-F(l))/(L_0(0)-L_0(l))]+1}\right]. \quad (27)$$

For the rectangular and circular lightwells,

$$F(0) = 1, \quad \text{and} \quad L_0(0) = L_u/2.$$

Therefore,

$$m=\frac{-1}{[2wL_u/(w+1)]\cdot[(1-F(l))/(L_u-2L_0(l))]+l}, \tag{26a}$$

and

$$b=\tfrac{1}{2}\left[1+\frac{1}{[2wL_u/l(w+1)]\cdot[(1-F(l))/(L_u-2L_0(l))]+1}\right]. \tag{27a}$$

For values $l>1.0$, the functions $F(l)$ and $2L_0(l)/L_u$ are small compared to unity and can be neglected. Therefore,

$$m\cong-\frac{(w+1)}{2w+l(w+1)}, \tag{26b}$$

and

$$b\cong\frac{w+l(w+1)}{2w+l(w+1)}. \tag{27b}$$

Note that both pairs of equations give values of m and b to define a linear function which passes through the point $\frac{1}{2}$ at $s=l/2$, i.e.,

$$ml/2+b=\tfrac{1}{2}. \tag{28}$$

Values of b, the intercept of $L(s)/L_u$, are tabulated in Table V as determined by the three methods considered. For the linear functions, m can be calculated from Eq. (28).

The results in Table V show that Eq. (27b) gives values more nearly checking the rather exact cinema integraph values than does Eq. (27a). This is particularly so for the 0.5-deep circular cylinder. Evidently the approximation made in deriving Eq. (27b) compensated for the earlier approximation of Eq. (24). This is most fortunate, since Eq. (27b) is very easy to use in computation. Accordingly, the solution of the integral equation for unit reflection factor is accurately given by a single straight-line approximation and the result is

$$\frac{L(s)}{L_u}=\frac{w+(w+1)(l-s)}{2w+l(w+1)}. \tag{21a}$$

TABLE IV.

s	$L(s)/L_u$ Simultaneous Equations	Cinema Integraph	Error
0	0.90051	0.8972	+0.00331
2	0.49999	0.5000	−0.00001
4	0.09948	0.1026	−0.00312

11. Extension of Simultaneous Equation Method for Reflection Factors Less than Unity

The simultaneous equation method can be extended to solve the integral equation when the reflection factor is less than unity by using three, four, or possibly five straight line segments. Since the interval over which a segment extends is arbitrary, it would be well to use wide intervals where the curve is expected to be fairly flat and narrow intervals where the curve is expected to bend sharply. The points used in forming the condition equations should be somewhat removed from the boundary of the intervals, and, therefore, the restrictive equations (such as the last equation of set (25)) would be avoided. It should be remembered that the formation of a particular condition equation requires

$$H(s_1)-\int_0^l K|t-s_1|\,H(t)dt=0 \tag{24a}$$

at the section s_1. Obviously this does not mean the correction function $H(s)$ is a minimum, nor does it mean $H(s_1)$ is zero.

The simultaneous equation method presented is somewhat similar to Fredholm's method of replacing an integral equation by a set of simultaneous equations.[8] However, Fredholm used line segments of zero slope as an approximating function. This, of course, halves the number of equations for a given number of intervals and simplifies the determination of the coefficients.

TABLE V.

w	Cross Section	b = Intercept of $L(s)/L_u$ From Eq. (27a)	From Eq. (27b)	Cinema Integraph
1	Rectangular	0.9004	0.9000	0.8972
2	Rectangular	0.8784	0.8750	0.8732
5	Rectangular	0.8578	0.8530	0.8549
1	Circular	0.7314	0.6667	0.6738

Nevertheless, the introduction of the slope greater than zero offers advantages in many cases. Further study is necessary to determine the scope and limitations of the method.

12. Conclusions

The solutions of the integral equations with the cinema integraph were very satisfactory in that results of practical value were obtained within a reasonable time. Also results of high accuracy were obtained by the perturbation method. These latter results were of value in determining the exact form of the luminosity function, and thereby suggested the mathematical proof for the general form of the function.

The value of the numerical results presented follows from the fact that the luminosity of the walls in a lightwell depends only upon the relative dimensions and not upon the absolute dimensions. Since the values of luminosity for reflection factors less than unity are not critical with depth, the same curves can be used for lightwells somewhat less deep than four units, or they can be extended slightly for deeper lightwells, and the accuracy will be good for practical calculations. By interpolation between the three cross sections considered and the different reflection factors, the luminosity in a variety of lightwells can be determined with an accuracy satisfactory for practical calculations.

Finally, as an important practical observation, a large increase in the lightwell luminosity can be obtained simply by constructing the side walls with a high reflection factor. Thereby the efficiency of a lightwell for the daylight illumination of lower floors can be greatly increased.

13. Acknowledgments

The writer expresses his gratitude to the Department of Electrical Engineering for the use of the cinema integraph, and to Professor Parry Moon for his continued interest and assistance in solving the problems. Professor H. C. Hottel of the Department of Chemical Engineering originally suggested the problem on the circular cylinder.

APRIL, 1941 J. O. S. A. VOLUME 31

Optical Reflection Factors of Acoustical Materials

Parry Moon
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received February 10, 1941)

1. Introduction

SOUND-ABSORBING materials are inherently rather poor reflectors of light because of their porous structure. Yet they are being used to an ever-increasing extent—being used, moreover, on ceilings, where light-reflecting properties are particularly important to the illuminating engineer. Thus the subject of the reflection factor of acoustical materials is of growing practical importance.

A number of investigations have been made of the light reflection from acoustical materials,[1,2] but none of the published results was obtained by use of a modern color analyzer. The color analyzer makes possible higher precision than can be obtained by other methods and also gives data on the chromaticity of the materials. The present paper tabulates values of reflection factor and chromaticity obtained by measuring 100 commercially available materials. The samples include by no means all the acoustical materials on the market, but it is believed that they give a representative picture of the various types available. Some samples are included which are usually classed as "insulating boards" rather than acoustical materials. All of these materials, however, have a rough, porous structure with some sound-absorbing properties; so it seems

[1] A. L. Powell and C. L. Dows, "Light reflection factors of acoustical materials," Trans. I.E.S. **25**, 882 (1930).
[2] J. M. Smith and L. S. Ickis, "Reflection factors of acoustical materials," Trans. I.E.S. **33**, 373 (1938).

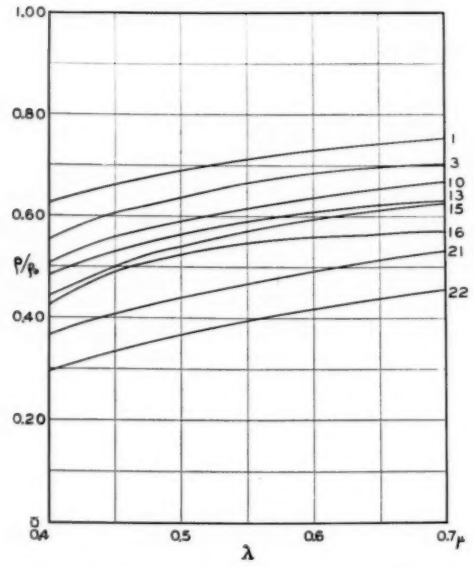


FIG. 1. Reflection from acoustical plasters and cast tiles. ρ = true spectral reflection factor of sample; ρ_0 = true reflection factor of MgO comparison surface.

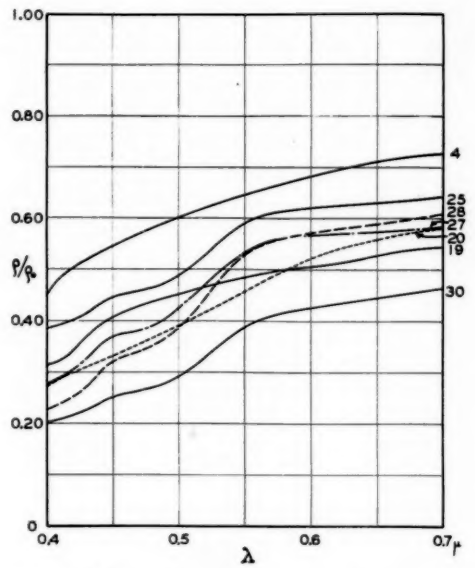


FIG. 2. Reflection from acoustical plasters and cast tiles. ρ = true spectral reflection factor of sample; ρ_0 = true reflection factor of MgO comparison surface.

reasonable to include them in the present group rather than with wood veneers, hard boards, etc., which will be considered in another paper.

2. Spectral Curves

Tests were made by means of the Hardy spectrophotometer at the Massachusetts Institute of Technology.[3] The instrument uses a beam of light that strikes the sample at almost normal incidence, and the reflected light is collected by an integrating sphere. In all the present tests, the integrating sphere was so arranged that any specularly reflected light that is present will be included in the results. The ordinates of the curves that are drawn by the machine represent $\rho(\lambda)/\rho_0(\lambda)$ where $\rho(\lambda)$ is the true spectral reflection factor of the sample at wave-length λ, and $\rho_0(\lambda)$ is the true reflection factor of the magnesium oxide comparison surface.

Results for a set of acoustical plasters and cast tiles are shown in Figs. 1 and 2. These materials would be classed by most people as "white" or "very light gray," yet the curves show that the reflection factors are disappointingly low. The curves are fairly uniform throughout the visible spectrum, with a tendency toward low values at

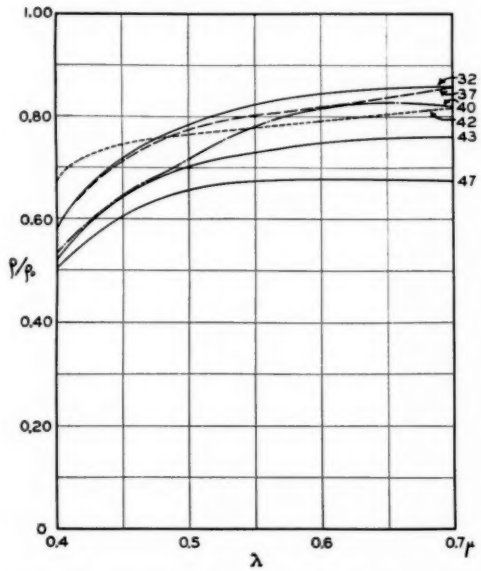


FIG. 3. Reflection from fibrous acoustical materials, white.

[3] I wish to thank especially Dr. S. Q. Duntley of the Massachusetts Institute of Technology Color Measurements Laboratory for his help in this investigation. I am also indebted to the manufacturers, who kindly supplied test samples.

the shorter wave-lengths. Figures 3 and 4 show what can be accomplished by painting the samples white. These are also commercial materials, tested as received from the manufacturers. The reflection factors are higher than with the samples of Figs. 1 and 2; but all materials reflect poorly in the blue end of the spectrum and are distinctly inferior to ordinary surfaces coated with modern white paints, which often exhibit 85 to 90 percent reflection factor.

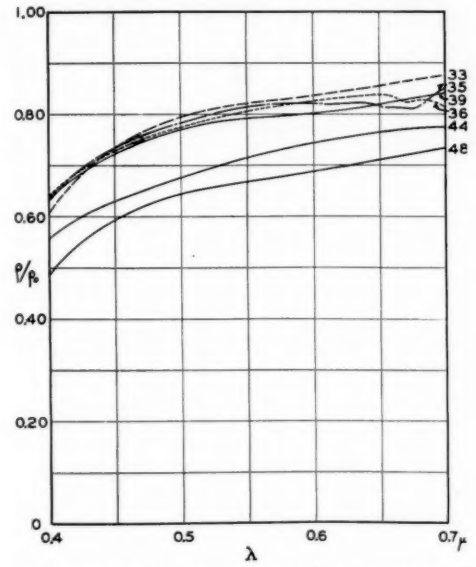


FIG. 4. Reflection from fibrous acoustical materials, white.

Figure 5 gives some results for fibrous materials, unpainted and having their natural straw color. The curves show the characteristic form usually associated with natural cellulose. Figures 6 and 7 show a slightly warmer tint obtained by dyeing or painting the acoustical materials. Some of the curves are fairly high, but with a decided dip near the middle of the spectrum. Figure 8 shows some pastel greens and blues, while Fig. 9 gives a few other colors.

The curves shown in the foregoing figures are designated by numbers which refer to Table I. Only about half of the samples of Table I are given on the graph sheets, but the omitted curves are so similar to those shown that it seems unnecessary to include them.

3. Tabulated Values

The illuminating engineer is ordinarily interested in values of total reflection factor obtained when the surface is illuminated by light from incandescent lamps. This reflection factor was

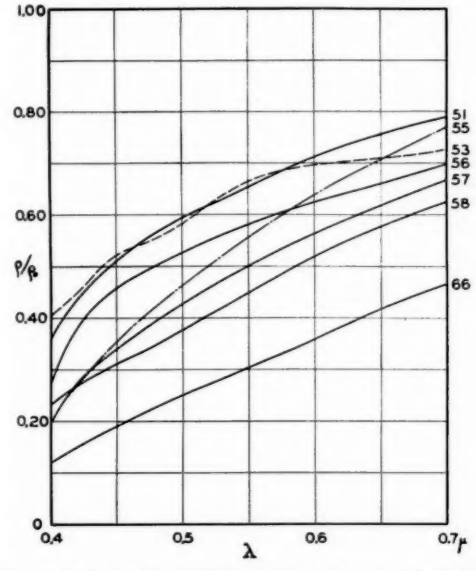


FIG. 5. Reflection from fibrous acoustical materials, natural color, unpainted.

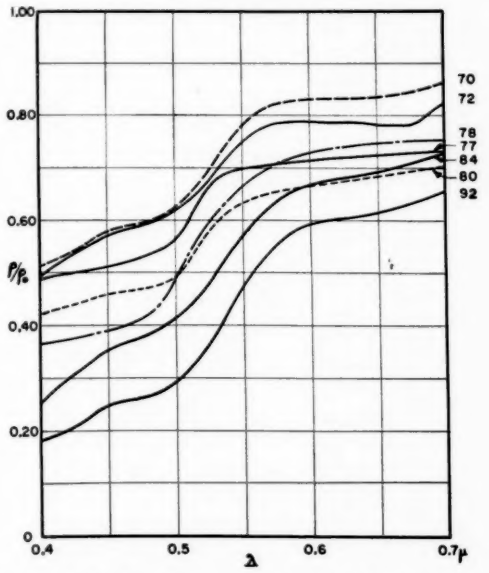


FIG. 6. Reflection from acoustical materials, ivory color.

TABLE I. *Reflection factors and trichromatic coefficients (for standard illuminant A).*

No.	NAME	MFGR.	APPROX. COLOR	x	y	ρ
	Acoustical Plasters and Cast Tiles, White					
1	Castacoustic, natural	13	Lt. gray	0.4547	0.4099	0.704
2	Akoustolith C, color No. 8	13	Lt. gray	0.4557	0.4116	0.678
3	Akoustolith, natural	13	Lt. gray	0.4566	0.4108	0.659
4	Calicel tile, painted white	6	White	0.4637	0.4126	0.643
5	Akoustolith C, color No. 3	13	Lt. gray	0.4572	0.4128	0.642
6	Acoustone D, white	30	Lt. gray	0.4568	0.4140	0.633
7	Acoustone F, white	30	Lt. gray	0.4566	0.4129	0.629
8	Akoustolith C, color No. 9	13	Lt. gray	0.4566	0.4110	0.629
9	Akoustolith, sprayed white	13	Lt. gray	0.4542	0.4118	0.617
10	Kalite D, natural white	7	Lt. gray	0.4580	0.4104	0.612
11	Akoustolith C, color No. 12	13	Lt. gray	0.4582	0.4112	0.608
12	Kalite A, natural white	7	Lt. gray	0.4548	0.4096	0.600
13	Plastacoustic, natural	13	Lt. gray	0.4572	0.4115	0.586
14	Akoustolith B, color No. 2	13	Lt. gray	0.4559	0.4121	0.573
15	Sabinite, white	30	Lt. gray	0.4597	0.4120	0.568
16	Akoustolith D, color No. 1	13	Lt. gray	0.4522	0.4170	0.549
17	Akoustolith C, color No. 10	13	Lt. gray	0.4549	0.4117	0.543
18	Akoustolith D, color No. 5	13	Lt. gray	0.4560	0.4119	0.534
19	Kalite H, No. 11, nat. gray	7	Lt. gray	0.4633	0.4134	0.483
20	Calicel tile	6	Natural	0.4807	0.4144	0.470
21	Kalite D, limestone	7	Lt. gray	0.4626	0.4110	0.469
22	Kalite A, limestone	7	Gray	0.4649	0.4117	0.397
	Acoustical Plasters, Warm Tints					
23	Kalite D, light buff	7	Buff	0.4737	0.4169	0.605
24	Kalite D, med. buff	7	Buff	0.4796	0.4189	0.594
25	Kalite A, light buff	7	Buff	0.4705	0.4165	0.581
26	Kalite A, med. buff	7	Buff	0.4845	0.4210	0.550
27	Akoustolith C, buff	13	Buff	0.4750	0.4204	0.526
28	Sabinite, ivory	30	Buff	0.4834	0.4231	0.522
29	Kalite D, sandstone	7	Buff	0.4762	0.4170	0.458
30	Kalite A, sandstone	7	Buff	0.4841	0.4189	0.387
	Tiles and Wall Boards, White					
31	Temcoustic	1	White	0.4579	0.4128	0.816
32	Acoustone D, painted white	30	White	0.4578	0.4130	0.812
33	Acousti-Celotex, 1322-L, No. 1 white	6	White	0.4558	0.4122	0.809
34	Acoustone F, painted white	30	White	0.4580	0.4133	0.802
35	Acoustilite	14	White	0.4538	0.4122	0.796
36	Econacoustic, white, sanded	22	White	0.4559	0.4115	0.792
37	Acousti-Celotex, type C-2, No. 3 white	6	White	0.4561	0.4126	0.790
38	Acoustex, white	22	White	0.4586	0.4117	0.789
39	Acousti-Celotex, 1322-L, casein finish	6	White	0.4543	0.4114	0.777
40	Limpet asbestos	16	White	0.4629	0.4142	0.773
41	Econacoustic, white (No. 14390)	22	White	0.4565	0.4111	0.768
42	Fir-Tex tile board, No. 300	12	White	0.4525	0.4090	0.768
43	Corkoustic	1	White	0.4572	0.4125	0.720
44	Acoustex, sunflex white	22	White	0.4595	0.4115	0.708
45	Econacoustic, white (No. 14391)	22	White	0.4623	0.4119	0.690
46	Insulating board, white glazecoat	15	White	0.4733	0.4141	0.690
47	Perfatone	30	White	0.4534	0.4131	0.662
48	Absorbex A, No. 1 white	6	White	0.4577	0.4117	0.660
49	Quietone, sanded, white	30	White	0.4674	0.4151	0.649

calculated from the spectral curves by the usual methods of mechanical integration.[4] Illuminant "*A*" (Planckian radiation at 2842°K, $C_2 = 14{,}320$) was considered in the calculations. Some of the integrations were performed automatically by the Hardy machine, which uses 100 selected ordinates. Other integrations were made by use of the experimental curves, a 30-ordinate method being used. The resulting values of ρ are given in Table I. These are true total reflection factors for 2842° radiation and are corrected for the reflection factor of the MgO comparison surface by multiplication of ρ/ρ_0 by $\rho_0 = 0.980$.

Integrations similar to those used in obtaining total reflection factor give the trichromatic coefficients x and y, Table I. The first column

[4] A Symposium on Color. I.E.S. Convention, Spring Lake, New Jersey, September, 1940.

TABLE I (*Continued*).

No.	NAME	MFGR.	APPROX. COLOR	x	y	ρ
	Tiles and Wall Boards, Natural					
50	Ins-Lite, fine screen surface	14	Natural	0.4738	0.4248	0.682
51	Acoustex	22	Natural	0.4714	0.4143	0.661
52	Ins-Lite, burlap surface	14	Natural	0.4780	0.4193	0.659
53	Acousti-Celotex, 4322-S, type C-2	6	Natural	0.4670	0.4164	0.655
54	Smoothcote	14	Natural	0.4808	0.4204	0.645
55	Econacoustic	22	Natural	0.4832	0.4227	0.585
56	Quietone, sanded surface	30	Natural	0.4684	0.4136	0.585
57	Quietone, brushed surface	30	Natural	0.4823	0.4169	0.515
58	Absorbex A	6, 25	Natural	0.4855	0.4146	0.465
59	Graylite, burlap surface	14	Brown (natural)	0.4855	0.4189	0.393
60	Graylite, fine screen surface	14	Brown (natural)	0.4879	0.4185	0.360
61	No. 83 tile board, smooth	6	Natural	0.4791	0.4152	0.352
62	No. 53 tile board, smooth	6	Natural	0.4782	0.4151	0.347
63	No. 53 tile board, tapestry finish	6	Natural	0.4899	0.4167	0.342
64	Celotex building board	6	Natural	0.4787	0.4140	0.338
65	No. 55 tile board, sanded	6	Natural	0.4775	0.4181	0.327
66	No. 55 tile board, tapestry finish	6	Natural	0.4911	0.4179	0.323
67	Studio board	6	Natural	0.5128	0.4105	0.292
68	No. 75 tile board, light brown, sanded	6	Lt. brown	0.5052	0.4144	0.275
69	No. 75 tile board, tapestry finish	6	Natural	0.5052	0.4138	0.265
	Tiles and Wall Boards, Buff					
70	Buff Satincote	14	Ivory	0.4685	0.4222	0.786
71	Temcoustic	1	Ivory	0.4754	0.4209	0.750
72	Acoustilite, buff	14	Ivory	0.4694	0.4166	0.735
73	Corkoustic	1	Ivory	0.4772	0.4217	0.720
74	No. 83 tile board, smooth, deep ivory	6	Buff	0.4692	0.4173	0.698
75	Acoustex, ivory	22	Buff	0.4764	0.4177	0.690
76	No. 83 building board, ivory	6	Ivory	0.4681	0.4156	0.685
77	Fir-Tex tile board No. 302	12	Ivory	0.4678	0.4195	0.679
78	Corkoustic	1	Buff	0.4857	0.4230	0.663
79	No. 83 tile board, smooth, deep ivory	6	Buff	0.4702	0.4180	0.640
80	Fir-Tex tile board No. 301	12	Cream	0.4729	0.4180	0.620
81	Econacoustic, cream	22	Cream	0.4889	0.4199	0.599
82	Quietone, sanded, cream	30	Cream	0.4785	0.4174	0.596
83	Acoustone D, ivory	30	Ivory	0.4747	0.4187	0.595
84	Quietone, sanded, buff	30	Buff	0.4914	0.4187	0.589
85	Acoustex, cream	22	Cream	0.4896	0.4206	0.581
86	Acoustone D, cream	30	Cream	0.4819	0.4225	0.577
87	Perfatone	30	Ivory	0.4645	0.4187	0.577
88	Insulation board	22	Cream	0.4873	0.4180	0.576
89	Celotex No. 83, building board, deep ivory	6	Buff	0.4751	0.4180	0.554
90	Acoustone F, ivory	30	Buff	0.4752	0.4196	0.551
91	Acoustone F, cream	30	Buff	0.4864	0.4235	0.531
92	Acoustex, buff	22	Buff	0.5062	0.4194	0.497
	Tiles and Wall Boards, Other Colors					
93	Gray Satincote	14	Lt. gray	0.4526	0.4097	0.689
94	Coral Satincote	14	Pink	0.4827	0.4142	0.684
95	Green Satincote	14	Lt. green	0.4299	0.4248	0.615
96	Quietone, blue	30	Lt. blue	0.4537	0.4182	0.576
97	Quietone, green	30	Lt. green	0.4600	0.4222	0.563
98	Quietone, rose	30	Pink	0.4958	0.4063	0.518
99	Acoustex, apple green	22	Lt. green	0.4603	0.4360	0.508
100	Acoustex, lt. gray	22	Gray	0.4677	0.4143	0.480

of the table indicates a designating number which agrees with the various curve sheets of this paper. The second column gives the manufacturer's specification of the material. The third column designates the manufacturer (see Table II), while the fourth column indicates very roughly the appearance of the sample.

A more precise specification of appearance is given by the values of x and y (Table I). If these values are plotted on the usual chromaticity diagram or color triangle, they form a small cluster, slightly to the right of the point representing standard Illuminant A. Figure 10 shows an enlargement of a small part of the complete

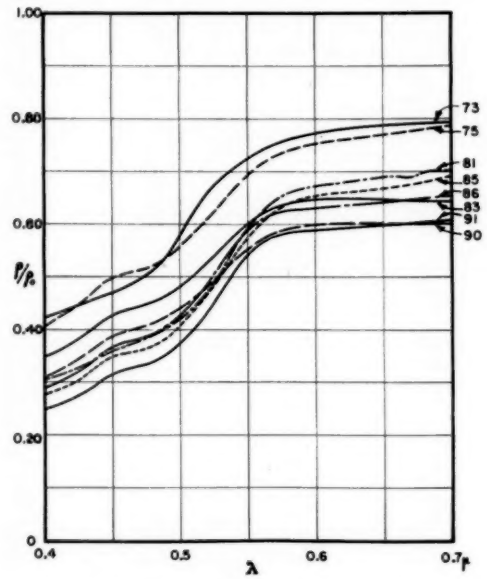


FIG. 7. Reflection from acoustical materials, buff.

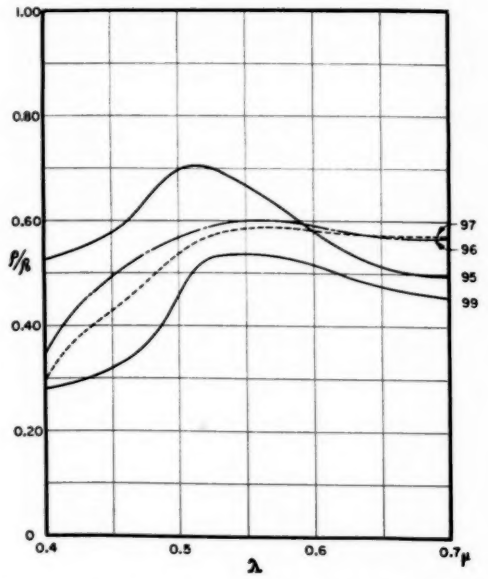


FIG. 8. Reflection from acoustical materials, blues and greens.

chromaticity diagram, Illuminant A ($x=0.4476$, $y=0.4075$) being somewhat to the left of the map shown. Except for unusual colored samples like Nos. 98 and 99, the materials are clustered near the Planckian locus and are all somewhat warmer in tone than the illuminant.

4. Effect of Various Factors

To determine the effect on the reflection factor ρ of changing the color of the incident light, two typical samples were used. One was a sample of natural fibrous material having a spectral curve that was almost a straight line, sloping downward to the left (as in Fig. 5). The other sample was painted white and had a typical curve like those of Figs. 3 and 4. Calculations were made for Planckian radiation ranging from 2000 to 10,000°K. Figure 11 shows that the natural-color sample gives a perfectly straight line when ρ is plotted against the reciprocal of the absolute temperature. A change from 2000 to 10,000° causes a change in reflection factor from 0.49 to 0.44. The white sample gives a curve that is slightly concave downward, and the total change in reflector factor is less than half as much as for the first sample. Some of

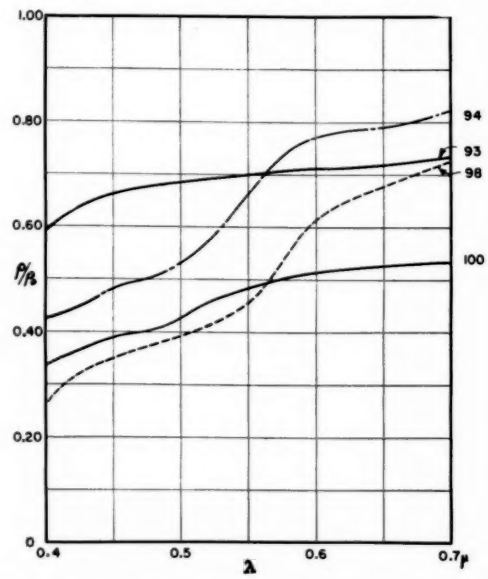


FIG. 9. Reflection from a few other acoustical materials.

the gray acoustical plasters would have shown even less change. These results are believed to be representative for acoustical materials and indicate that a change from incandescent lamps

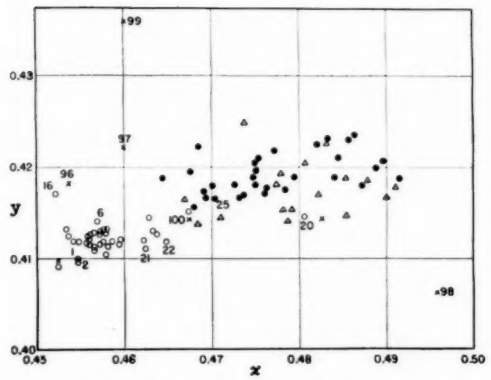


FIG. 10. Chromaticity diagram. ○ white samples; ● buff; △ natural; X others.

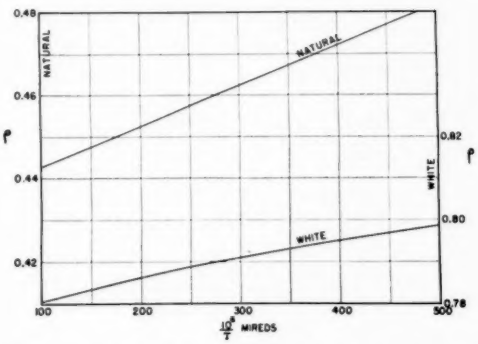


FIG. 11. Effect of the color of the illuminant on the reflection factor. Planckian radiation, temperature varied from 2000 to 10,000° K (500 to 100 mireds).

to daylight fluorescent lamps or to the various phases of daylight will cause ordinarily only 0.02 to 0.03 change in the reflection factor of these materials.

The data of Table I are believed to be accurate representations of the particular samples tested. It must be realized, of course, that there are variations in any manufactured product, also variations in the composition of the paint and in the way in which it is applied. Reflection factors obtained on these materials in actual use will undoubtedly be lower than the test values because of the collection of dust and dirt. All samples used in this research were new and clean, and no attempt was made to evaluate quantitatively the effect of age.

The method of test also has some effect on the results.[5] A change in the angle of incidence

TABLE II. *Manufacturers.*

DESIGNATION NO.	NAME
1	Armstrong Cork Co.
2	Asher & Boretz, Inc.
3	Bigelow-Sanford Carpet Co., Inc.
4	E. L. Bruce Co.
5	Cambridge Tile Mfg. Co.
6	Celotex Corp.
7	Certain-teed Products Corp.
8	Columbia Mills, Inc.
9	Columbus Coated Fabrics Corp.
10	Congoleum-Nairn, Inc.
11	E. I. du Pont de Nemours & Co.
12	Fir-Tex Insulating Board Co.
13	R. Guastavino Co.
14	Insulite Co.
15	Johns-Manville
16	Keasbey and Mattison Co.
17	David E. Kennedy, Inc.
18	Libbey Owens Ford Glass Co.
19	Marsh Wall Products, Inc.
20	Masonite Corp.
21	Mosaic Tile Co.
22	National Gypsum Co.
23	National Tile Co.
24	Natural Slate Blackboard Co.
25	Northwest Magnesite Co.
26	Paraffine Companies, Inc.
27	Pittsburgh Plate Glass Co.
28	Sherwin-Williams Co.
29	Tile-Tex Co.
30	U. S. Gypsum Co.
31	Uvalde Rock Asphalt Co.
32	Westinghouse Electric and Mfg. Co.

TABLE III. *Comparison of results.*

MATERIAL	MFGR.	REFLECTION FACTOR S & I	REFLECTION FACTOR MOON
Absorbex, unpainted	6	0.39	0.465
white	6	0.57–0.63	0.660
Acousti-Celotex mineral tile, painted white	6	0.80	0.809
Acoustex, natural	22	0.61	0.661
white	22	0.73	0.789
Acoustone D, cream	30	0.52	0.577
ivory	30	0.57	0.595
white	30	0.65	0.633
Corkoustic, buff	1	0.58	0.663
ivory	1	0.65	0.720
white	1	0.73	0.720
Castacoustic C, natural	13	0.56	0.704
Perfatone, white	30	0.73	0.662
ivory	30	0.63	0.577
Plastacoustic, natural	13	0.54	0.586
Quietone, brushed surface, natural	30	0.53	0.515
sanded, natural	30	0.57	0.585
sanded, white	30	0.68	0.649
Sabinite, white	30	0.53	0.568
ivory	30	0.53	0.522
Temcoustic, white	1	0.76	0.816
ivory	1	0.75	0.750

[5] H. J. McNicholas, "Absolute methods in reflectometry," Bur. Stand. J. Research 1, 29 (1928).

of the light used in the tests affects the reflection factor, as does a change in the method of collecting the light reflected from the sample. An attempt was made to compare the reflection factors obtained in this research with the results of previous tests. Acoustical materials, however, have changed so much since 1930 that no comparison was possible with the values of Powell and Dows.[1] Even with the results of Smith and Ickis[2] there is difficulty in knowing whether the materials having the same name are actually similar or have been modified since 1938. Table III shows the comparison and indicates that the present results are usually somewhat higher than those of Smith and Ickis. Whether this is a difference in the samples or is caused by the completely different test methods and apparatus is not clear.

5. Conclusions

Reflection factors and trichromatic coefficients are obtained for 100 samples of acoustical materials and insulating boards. When standard Illuminant *A* is used, the values of reflection factor range from 0.265 to 0.816. The chromaticity range is small and even the white samples are found to be somewhat warm in tone. Reflection factors vary only a few percent when the incident Planckian radiation is varied from 2000 to 10,000°K, showing that ordinary variations in color of the illuminant have almost negligible effect on the reflection factor of acoustical materials.

No attempt was made to determine how rapidly the reflecting properties would deteriorate when the acoustical materials were exposed to ordinary conditions in rooms. A glance at the materials, however, is sufficient to show that the surfaces are among the worst dirt-catchers that the ingenuity of man has yet devised. The only type of acoustical material that can be wiped clean with a cloth and that can be repainted as often as desired without loss of acoustical qualities appears to be the perforated metal tile with absorbing pad. Moreover, this type of material has almost perfect sound absorption, despite the fact that the total area of the holes constitutes only a small percentage of the surface area. Some of the properties of these materials[6] are summarized in Table IV.

Table IV. *Perforated metal tiles.*

Name	Noise Reduction Coef.	Perforations: Diameter	Perforations: No. per sq.ft.	Perforations: Percent of Total Area
Acousteel	0.85	0.094 in.	2054	10
Perfatone	0.85	0.073	4608	13
Sanacoustic	0.85	0.068	4608	12
Ordinary fibrous acoustical material	0.60	—	—	—

Table IV shows that the holes reduce the reflection factor to approximately 0.88 of its value for a continuous painted surface, or

Reflection Factor of Paint	Calculated ρ of Tiles
0.90	0.79
0.85	0.75
0.80	0.70.

By using a good white paint, therefore, over-all reflection factors of more than 70 percent can be obtained for a ceiling of perforated metal tiles. The combination of high acoustical effectiveness and freedom from deterioration of reflection factor means that the perforated metal tile merits serious consideration.

[6] Acoustical Materials Association, Chicago, Illinois, Bulletin No. 7 (1940).

The Geometry of Stereoscopic Projection

JOHN T. RULE
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received January 23, 1941)

RECENT improvements in the stereoscopic art have greatly increased the importance of stereoscopic projection. Literature on the subject is either nonexistent or fragmentary. A demand has consequently arisen for a unifying article on the theory and practice of such projection. Thus, much of the material in this article is not new but only correlated into a consistent pattern.

DIFFERENCE BETWEEN TWO-AND THREE-DIMENSIONAL PHOTOGRAPHY

Each point in a single two-dimensional photograph may represent any point in space on the ray from the eye through that point. The mind subjectively decides the distance along this ray to the point which is supposed to be represented. This decision seems to be achieved by a summation, in the light of experience, of all the evidence presented by the total picture. Geometrically, an infinite number of judgments are permissible. A small photograph of a sphere may be interpreted as a small sphere at a short distance or a large sphere at a proportionally greater distance. If the same photograph be enlarged any number of times and presented to the eye so that it subtends the same monocular angle, exactly the same interpretations are permissible. We may thus say that, geometrically, a single photograph is only fixed in direction.

In a stereograph the two images of a single point might each, taken singly, represent any point on the ray from the proper eye through that point. Taken together, however, properly fused by the two eyes, they may represent only the single point where the two rays from the two eyes intersect in space. The stereograph of a single point is thus fixed both in direction and in distance. Thus an entire stereograph is geometrically fixed in direction, distance, shape and size. It is susceptible of only one interpretation instead of an infinite number.

For the purposes of our analysis the above may be considered a true statement. It is a fact, however, that eyestrain, evidence conflicting with experience, and other subjective, little understood factors, cause the brain to misinterpret stereoscopic images in varying degrees. This is a very complex and obscure subject which warrants a great deal of further research. These variations, however, do not alter the basic fact that a single fixed image is established.

One other important difference between a stereograph and a flat photograph exists. In viewing a flat photograph the eyes are focused for the distance to the photograph. The two eyes in looking at any point, since there is only one view of it, are also converged for the distance to the photograph. Thus the focus and convergence are for the same distance. This is in accord with normal space vision.

In viewing a stereograph, the eyes are always focused for the distance to the stereograph. (We shall not consider any form of lens viewing in which this is not the case.) They are, however, converged for the distance to the intersection of the two rays from the eyes through the two views which represent the single point at which they are looking. The convergence thus changes for different points in the image while the focus remains constant. Thus the accommodation-convergence ratio normal to space vision is violated. This seems to be a source of eyestrain which must be carefully considered in all stereoscopic projection.

PARALLEL PROJECTION

The simplest type of projection is parallel projection. By this is meant that the axes of the cameras are parallel in taking and the pictures are projected so that the two infinity points are the distance of the human eyes, called the interocular distance, apart on the screen. By infinity points are meant the images of a point at infinity directly in front of the camera. If the axes of the two projectors are parallel and interocular apart, the latter condition is achieved regardless of the size of the images on the screen.

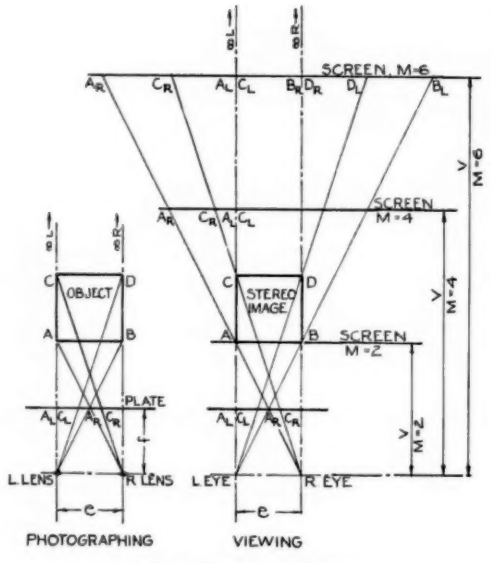


FIG. 1. Parallel projection.

If the axes of the two projectors are greater than the human interocular apart a suitable toe-in must be maintained to make the infinity points interocular apart on the screen. This will be classified as parallel projection though it introduces a keystoning effect not present in the ideal case.

If vectographs are used in which the two views are superimposed on a single film using a single projector, the original vectograph should be so registered that the separation between infinity points on the vectograph multiplied by the magnification from the film to the screen equals the human interocular. This will correctly locate the infinity points on the screen. After the vectograph has been made there is only one theoretically correct size at which it can be shown to achieve this exactly.

Figure 1 illustrates the basic features of parallel projection when both the camera axes and the projector axes are interocular apart. For clarity of explanation, the plate of the camera is shown in front of the lenses, rather than behind them, a distance equal to f, the focal length of the camera.

It will be seen that, under the above conditions, if the eyes replace the camera lenses and the screen replaces the plate, the two images on the screen being unmagnified, the eyes will see a stereoscopic image of the same size, shape and location as the original space scene. If now the screen be removed to a greater distance and the magnification around the infinity points be increased proportionally so that the two pictures subtend exactly the same monocular cones, exactly the same stereoscopic image will be seen, the paired rays intersecting at exactly the same points in space as previously.

Thus, with parallel projection, there will always be a position from which the resultant stereograph may be viewed to yield a stereoscopic image of the proper size, shape and location. If the two pictures are enlarged n times, the proper viewing distance will be closely equal to n times the focal length of the camera.

Expressed mathematically, $V = Mf$ where V = proper viewing distance; M = magnification, or the screen width divided by the film width; f = focal length of the camera.

It is assumed that the camera is focused at infinity as the error introduced by this assumption is negligible compared to normal errors in viewing locations.

Certain further important observations may be made from Fig. 1.

Some frontal plane in the original scene will appear to lie in the plane of the screen wherever the screen may be placed. The proper viewing distance is then equal to the distance from the camera lenses to that frontal plane in the original scene. Any object in front or behind this plane will appear a corresponding distance in front or behind the screen. This plane always exists and is that plane for which the two views of all points in it appear superimposed on the screen.

As the magnification increases, the distance from the screen to the proper viewing point also increases,—in effect the screen recedes into the scene, which remains stationary. This means that the near object, as magnification increases, appears to be farther and farther in front of the screen with a consequent greater breakdown of the normal accommodation-convergence ratio. This may cause either a great eyestrain or a complete inability of the observer to fuse the two pictures.

In Fig. 1, for example, the front edge of the cube shown appears in the plane of the screen when $M = 2$, but one-third of the way from the

eyes to the screen when $M=6$. This ratio will remain fixed regardless of the position of the observer.

This immediately points out the necessity of establishing a criterion for permissible variation of image location relative to the screen. Eyestrain is demonstrably present for images relatively far in front of the screen. For our purposes whether the accommodation-convergence breakdown is the cause of this or not is unimportant. Not enough experimentation has been done to establish the proper terms in which to express such a criterion. We shall arbitrarily use the percentage of the distance from the eyes to the screen of the front of the image.

Some help can be obtained by considering the variations common to normal vision. Most pairs of eyes can both see over an angle of about 90 degrees without motion of the head. A point 45° to either side of the center, which yields the same convergence angle as a point straight ahead, is distant from the eyes about 70 percent of the distance to the point straight ahead. Thus about a 30 percent variation of the accommodation-convergence ratio is encountered in normal vision.

Tests made in establishing the conditions under which the Chrysler movie at the 1940 New York World's Fair were to be shown seemed to indicate that objects may be at least halfway from the screen to the observer for the shortest viewing distance liable to be encountered in the usual hall, without causing noticeable strain to the great majority. This ratio, however, is extremely arbitrary and unquestionably increases for longer viewing distances. It will be used arbitrarily here in the absence of more exhaustive study. There is no question but that the safest and most satisfactory procedure is to have the near point of the image in the plane of the screen.

It can be seen in any case that, as magnification increases, straight parallel projection imposes a severe limitation with respect to the permissible distance to the near object in photographing. In fact, under practical conditions it effectively eliminates the possibility of close-ups.

For example: a standard 35-mm picture, effective width 0.825 inch, taken with a 35-mm lens and projected so that the picture is 15 feet wide, not an excessive screen width for movies, will have a proper viewing distance of slightly more than 25 feet. Thus any object less than 25 feet from the camera would appear in front of the screen and no object could be closer than 12.5 feet if we are to make our image limit halfway from the screen to the eyes.

It is also to be noted that the ideal viewing distances for most standard cameras are such that, under usual projection conditions, the majority of the audience is considerably in back of the best viewing point. This leads to an increased depth which is much more noticeable on near objects. This is highly objectionable in the usual close-ups of known shaped objects such as people or room interiors.

Stereoscopic Magnification

There are two possible methods of eliminating the above difficulties.

The first is to use a distance between camera lenses of less than interocular. Figure 2 illustrates this method.

Any scene in space photographed with a given distance between lenses, T, will yield exactly the same two views as a similar scene n times as large at n times the distance with a distance between lenses of n times T. The only difference will be the distance between the two views. Thus a scene photographed with a distance between camera

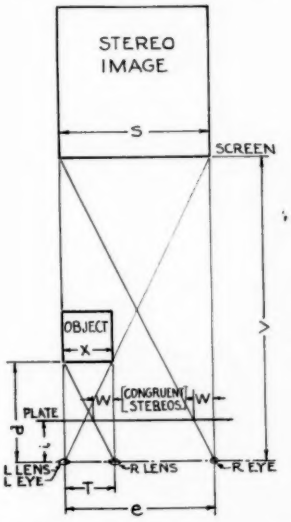


FIG. 2. Principle of stereoscopic magnification.

lenses, T, of less than interocular, if projected parallel with infinity points interocular apart, will appear the same as, and will be interpreted as, a similarly shaped, larger scene, farther away. This makes it possible by the choice of a suitable lens separation to maintain the shape of any scene while increasing its size and its distance. The latter fact permits the choice of the frontal plane in the scene which it is desired to place in the plane of the screen.

In Fig. 2 assume that the front face of the cube shown is to appear in the plane of the screen. w = width of picture on film; s = width of picture on screen; T = lens separation of camera; e = human interocular; f = focal length of camera; d = distance from camera lenses to plane in scene which is to coincide with the plane of the screen.

From similar triangles:—

$$i = wd/x, \tag{a}$$

from the simple lens formula

$$\frac{1}{i} + \frac{1}{d} = \frac{1}{f}$$

or

$$i = df/(d-f), \tag{b}$$

from (a) and (b)

$$x = w(d-f)/f. \tag{c}$$

Also, since the two figures are similar,

$$x = sT/e, \tag{d}$$

eliminating x from (c) and (d) and solving for T we have

$$T = we(d-f)/sf. \tag{1}$$

It is to be remembered that this is a formula for determining the separation between camera lenses in order to place any desired frontal plane in the plane of the screen while maintaining parallel projection. In general, it will be desired to place the near object in this plane as the purpose of the formula is usually to eliminate images in front of the screen in order to cause the least eyestrain.

Unless very near objects are photographed or long focal length lenses used, the value of f with respect to d is negligible and this formula may be effectively simplified to

$$T = wed/sf. \tag{1a}$$

From similar triangles the correct viewing distance then becomes

$$V = ed/T \quad \text{or} \quad V = sf/W.$$

Since most cameras have a fixed interocular such that $T = e$, it becomes of interest to know the maximum size screen that can be used with such cameras without the image of the near object coming in front of the screen or, given the size of screen, to determine the minimum distance to an object to prevent its image from being in front of the screen. When $T = e$ the formula reduces to

$$Sf/W = d. \tag{e}$$

Figure 3 shows a graph for this formula for various standard focal lengths.

It can be seen that a variable lens separation system enables us to photograph very small objects and show them at a large stereoscopic image size, not to be confused with screen size, still maintaining them near the plane of the screen. It is, in essence, stereoscopic magnification and will be so called. In general, the procedure is satisfactory but certain limitations should be

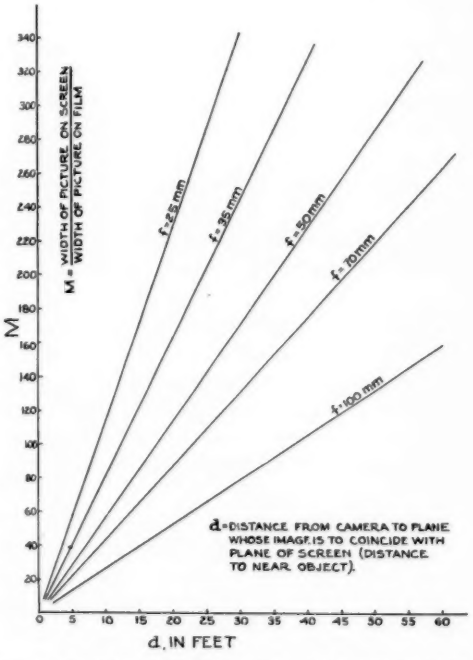

M = WIDTH OF PICTURE ON SCREEN / WIDTH OF PICTURE ON FILM
M
f=25 mm
f=35 mm
f=50 mm
f=70 mm
f=100 mm
d=DISTANCE FROM CAMERA TO PLANE WHOSE IMAGE IS TO COINCIDE WITH PLANE OF SCREEN (DISTANCE TO NEAR OBJECT).
d, IN FEET

FIG. 3. Parallel projection chart for placing near object in plane of screen for standard interocular cameras.

noted and guarded against in practice. This type of magnification differs from the magnification of a single photograph. A single photograph of a human head shown nearly filling a 15-foot screen may be interpreted as a normal sized head at a short distance. A stereoscopic magnification of the same head to place the stereoscopic image at the screen fixes it as a giant 15-foot head.

Interpretations of this phenomenon of giantism seem to differ. With some observers if the true sizes of the objects photographed are known the giantism may pass unnoticed unless the attention is directed to it. With others it may be immediately apparent and very annoying.

Two further facts are to be noted. If the size of the object is not known to the observer, he will interpret it as truly of the giant size. A small model of a known object, such as a locomotive, may be made to appear the size of the true object, the real locomotive, by the choice of a suitable lens separation.

A second objection to this system is much more serious when great magnifications are involved. This has to do with the limits of stereoscopic vision. The difference in convergence angle between the front and back of a small object at a small distance is greater than the difference between the front and back of a similar larger object at a proportionally greater distance. Since the sense of depth is less for smaller differences of convergence angles, there is a definite loss of "depth sense," or perception of depth—as distinguished from depth itself—in a stereoscopic magnification. If the magnification is great enough, the difference in convergence angles may be reduced below the ability of the eye to resolve them and a flat object results. For example, if a one-inch cube is photographed from two feet with a suitable lens separation to magnify it to fill a 15-foot screen it will appear as a 15-foot cube if viewed from 360 feet. Now at 360 feet the difference in convergence angle between the front and back of such a cube will have become so small that the eye cannot resolve it and the stereograph will appear perfectly flat.

If the viewing distance is reduced from 360 feet, even though the eye may now resolve the difference in depth from the front to the back of the cube, this depth becomes steadily less as the screen is approached. Thus at a viewing distance of 60 feet the supposed 15-foot cube would appear to be only $2\frac{1}{2}$ feet deep.

For this reason where great magnification has been employed the results have been disappointing, the stereoscopic effect either being completely lost or so slight as to be unsatisfactory. No specific limits of permissible magnification can be given as it is ultimately a question of artistic judgment. An experienced photographer will tend arbitrarily to increase the indicated value of T when he is in doubt.

If a lens separation greater than the human interocular is used the converse to the above is true, that is, the scene becomes nearer and smaller while the depth sense is increased.

The breakdown of the accommodation-convergence ratio for images back of the screen never seems serious if the observer does not approach the screen too closely. Eyes focused for twenty feet or more have very little difficulty converging from this distance to infinity. Consequently, it is seldom important to bring the image of an object in back of the screen forward to the plane of the screen unless something else is gained. Size is definitely lost in such a procedure and this may produce a distracting miniature effect. It is, consequently, generally advisable to use a lens separation equal to interocular where the distance to the near object indicates a lens separation greater than this. This places the entire image in back of the screen but avoids producing the effect of a small model. In truly distant scenes where no depth would be apparent without an increased lens separation, such as mountain pictures or aerial photographs, it may be desirable to sacrifice size to obtain depth sense.

For clarity of understanding it can be roughly stated that the depth sense varies inversely as the stereoscopic magnification. A choice of one must be made at the sacrifice of the other.

Nonparallel Projection

The other method of adjusting the image to the screen is to increase the separation between the infinity points of the two pictures, that is, to depart from parallel projection. This has the effect of moving the entire stereoscopic image away from the observer. It has serious drawbacks. It introduces distortions of some magnitude chiefly in the direction of increased depth. It

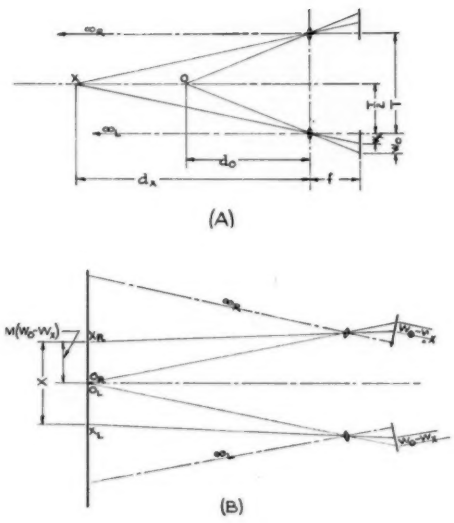


FIG. 4.

tends to cancel partially the effect of a close-up and it is limited in depth in that the separation of the two views of a point at the maximum depth in the scene must not be more than standard interocular. This strictly limits the permissible depth of the scene which grows less the greater the magnification and the greater the separation.

However, it is frequently necessary to make use of this method chiefly because of an inability to use a variable lens separation. Scenes containing motion cannot be photographed by shifting a single camera and variable-lens-separation, double cameras are difficult to design for small separations and at present not generally available. The problem therefore becomes:—Given a camera of fixed lens separation or of fixed minimum separation and a predetermined size of screen, what is the maximum depth of scene permissible?

We shall arbitrarily assume that no separation of the two views of a single point on the screen shall exceed the human interocular. Though some divergence can be handled by a normal pair of eyes, certainly they should not be required to do so. Furthermore, even this assumption means that we are theoretically going to remove the far point in the scene to infinity in the image. Though this may be a gross theoretical distortion the subjective interpretation is usually much less so. It is certainly not too bad a compromise when the focus of attention is some near object.

Figure 4(a) shows the taking system in which 0 is a point on the center line which is to appear in the plane of the screen and x any more distant point.

From similar triangles

$$W_0 = fT/2d_0, \tag{f}$$

$$W_x = fT/2d_x. \tag{g}$$

Figure 4(b) shows the projection system.

$$x = 2M(W_0 - W_x), \tag{h}$$

whence

$$x = MfT(1/d_0 - 1/d_x). \tag{k}$$

The maximum value of x is e when d_x becomes $d_{\max}$. Under these conditions

$$e = MfT(1/d_0 - 1/d_{\max}), \tag{m}$$

whence

$$d_{\max} = d_0 MfT/(MfT - ed_0). \tag{2}$$

For a standard stereoscopic camera in which $T = e$ this formula becomes

$$d_{\max} = d_0 Mf/(Mf - d_0). \tag{2a}$$

This formula is useful for two purposes. Given a fixed camera and the size of screen to be used, the maximum depth of scene may be determined when placing any point in the plane of the screen. Conversely, if pictures have already been taken or if the depth of scene is fixed by external circumstances, the maximum permissible size of screen may be determined. Formula (2a) lends itself readily to a very useful nomograph. This is shown as Fig. 5. If $d_{\max}$ is infinity the formula reduces to formula (e) and Fig. 3 becomes applicable.

This formula may also be used to determine the maximum depth of scene to permit the image to extend from halfway between the observer and the screen to infinity since we have set these limits as determining our practical maximum image range. If the scene is to start halfway to the screen the d_0 in the formula becomes d_N, or the distance to the near object. The images of this point on the screen will be $-e$ apart, that is, crossed over. The images of the most distant

point are $+e$ apart so that $x=2e$ or

$$2e=MfT(1/d_N-1/d_{\max})$$

or

$$d_{\max}=d_N MfT/(MfT-2ed_N). \qquad (3)$$

It is to be remembered that the use of all these compromise formulae abandons any effort to maintain a properly shaped image. As a matter of practice, however, in scenic work the enormous exaggerations of depth which occur by photographing a scene of limited depth and placing the far point at infinity in the image, are much less annoying than theory would indicate. The danger in its use lies in the distortion of rather geometric shapes such as rectangular rooms, table tops, etc. Experience or trial efforts are the only way of judging this.

Two methods of obtaining the increased separation may be employed. The best of these is to toe-in the axes of the camera if this is possible. This is a help in the problem of adequate framing to be discussed later. If it is not possible to toe-in the camera, the projectors may be toed-out or their separation increased. The toe-in of the camera is valuable as a framing device and it centers the scene, consequently reducing the non-stereoscopic area of the picture. However, unless the toe-in of the camera and the projectors are equal, a keystoning effect results, that is, the left side of the left picture will have a greater height than the left side of the right picture and vice versa. This makes complete vertical alignment impossible. Though the eyes are capable of handling a few degrees of vertical error, some eyestrain is necessarily introduced. Experience has shown that the effect is not important except for very near scenes, which should not be attempted unless a reduced lens separation is possible.

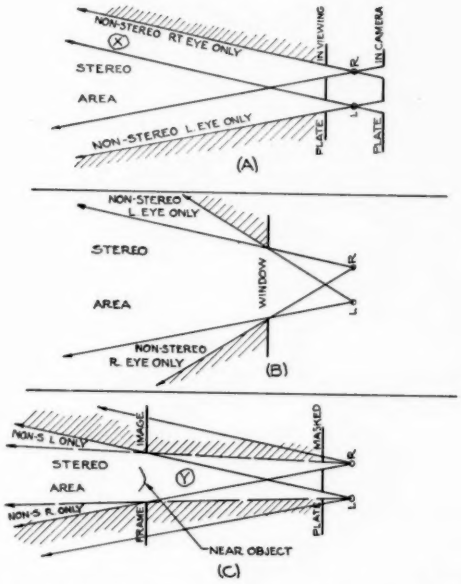


FIG. 6. (A) Viewing unframed stereo; (B) normal vision through a window; (C) viewing properly framed stereo.

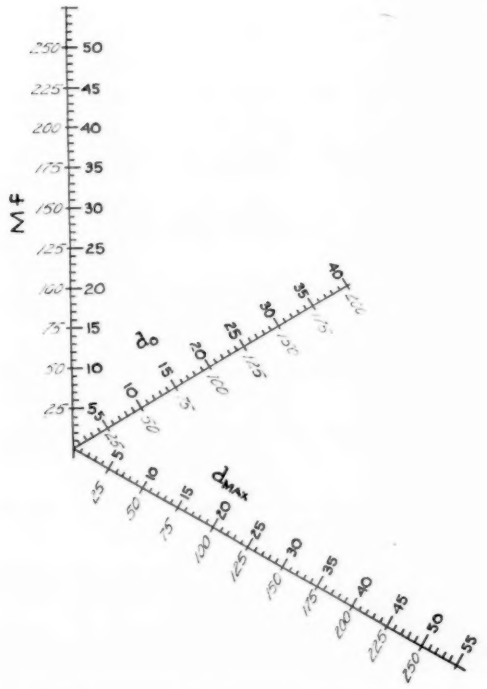


FIG. 5. Nomograph for determining maximum depth of scene when near object is to be placed in plane of screen by increasing separation between views. For standard interocular cameras only. Formula: $d_{\max}=d_0Mf/(Mf-d_0)$. d_0=distance to near object. Instructions: Mf, d_0 and $d_{\max}$ must be collinear. Note: $Mf=S\cdot f/w$. S=width of screen; w=width of picture on film; f=focal length.

The use of a toe-in for the camera will slightly alter formula (h). The proper formula is

$$x=4MfT(d_x-d_0)/(T^2+4d_xd_0). \qquad (n)$$

This will not be derived here as the quantity T^2 is negligible compared to d_x and d_0 and its

elimination reduces formula (n) to formula (k).

The following general procedure for taking and projecting stereographs will serve to summarize the discussion to this point.

First determine the maximum size of screen that is to be used in projection. In formula (1) if a fixed lens separation camera is used, substitute the distance between lenses for T. Then solve for d.

The resultant value of d is the control distance to the near object. For all distances greater than this pictures may be taken and projected parallel without any image being in front of the screen and with no viewing difficulties. The great majority of pictures usually fall in this category.

With a fixed lens separation camera it is advisable to avoid taking pictures of objects nearer than the above value of d. If, however, such pictures are to be taken, solve formula (2) or use Fig. 6 to determine the value of $d_{\max}$ and, if possible, limit the depth of scene to this value, then in projecting adjust the projectors so that the image of the near object is fused at the screen. If it is impossible to limit the depth of the scene the only alternative is to adjust the separation of the images of the far point in the scene to interocular separation on the screen, allowing part of the scene to come forward from the screen. Formula (3) may be used to determine the utmost that can be done and still have a viewable picture.

If a variable lens separation camera is used, or if a single camera is to be shifted solve formula (1) for T and use this value up to $2\frac{1}{2}$ inches and project parallel. For values of T greater than $2\frac{1}{2}$ inches use $2\frac{1}{2}$ inches.

Remember that, in reality, no rigid limits exist. Violations of the formulated technique given here lead to divergence, too great a breakdown of the accommodation-convergence ratio, bad image distortion, or combinations of these.

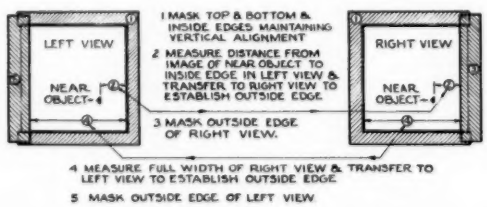

1 MASK TOP & BOTTOM & INSIDE EDGES MAINTAINING VERTICAL ALIGNMENT
2 MEASURE DISTANCE FROM IMAGE OF NEAR OBJECT TO INSIDE EDGE IN LEFT VIEW & TRANSFER TO RIGHT VIEW TO ESTABLISH OUTSIDE EDGE
LEFT VIEW
NEAR OBJECT
RIGHT VIEW
NEAR OBJECT
3 MASK OUTSIDE EDGE OF RIGHT VIEW.
4 MEASURE FULL WIDTH OF RIGHT VIEW & TRANSFER TO LEFT VIEW TO ESTABLISH OUTSIDE EDGE
5 MASK OUTSIDE EDGE OF LEFT VIEW

FIG. 7. Steps in framing a stereograph.

Now some divergence can be handled by normal eyes, a rather wide breakdown of the accommodation-convergence ratio is at least viewable, and distortions are inevitably present except under ideal conditions for the single individual who may be located at exactly the proper viewing point. Only experience, however, can dictate the extent of the liberties that may be taken with the ideal procedure.

FRAMING

The proper framing of stereographs has been partially discussed in many articles. It is of great importance in eliminating eyestrain, preventing impossible interpretations and enhancing the stereoscopic effect. A stereograph should be seen through a window. This window should be sharp, and in general, fusible when the eyes are looking at any point in the scene near the edges. The technique of establishing this window needs to be thoroughly understood.

The stereoscopic camera with axes parallel takes two pictures covering different areas as shown in Fig. 6(a). It will be seen that the two nonstereoscopic areas are seen by the wrong eyes as compared with the nonstereoscopic areas when looking through a window as shown in Fig. 6(b). Stated specifically, an unframed stereograph shows farther to the left in the left-eye picture and farther to the right in the right-eye picture, while in looking through a window the left eye sees farther to the right and the right eye farther to the left. To frame a stereograph properly this latter condition must be exactly duplicated. Figure 6(c) shows how the stereograph in Fig. 6(a) is masked to achieve this.

The results of failing to frame properly should be pointed out. Consider the object x, Fig. 6(a). This appears only in the right-eye picture. In the left-eye picture it is beyond the edge and, consequently, if the edges of the picture are to be interpreted as a window it should not be visible to the right eye at all. The resultant interpretation is confusing as the mind is presented with an entirely unnatural situation. The object appears double, the edges in the two eyes not being fused. This causes a retinal rivalry which distracts the attention from the picture.

In Fig. 6(c) consider an imaginary object at y. Here the stereograph has been framed as shown.

The object is in the stereo area but it appears in front of the window. If it lies near the edge of the picture and sufficiently forward from the frame the eyes cannot converge on both at once and a doubling of the frame occurs when the object is being viewed. This may be eliminated if the object is in the center of the picture. In this case the scene from the frame to the object protrudes through the frame and is cut off below and above by the frame. Thus, if the object is resting on the ground, the ground seems to come through the frame without visible means of support, or if the ground is not visible the object itself seems to be cut off and suspended in air without means of support. The same is true if the object comes down from the top. If it extends beyond the sides it seems to be only a part of the true object unnaturally cut off in space. These effects are very real and annoying ones and should consequently be avoided.

We can thus say that the frame of a stereograph should appear as a window in front of all objects in the scene yet sufficiently close to the scene at the edges to prevent doubling due to change in convergence. The only permissible exception to this is the case of an object near the center of the picture which may be forward from the frame provided it is completely visible and supported by an extension through the frame. Thus a man's arm or a piece of a machine may come through the frame, provided the man himself or the machine itself is standing behind the frame.

The simplest method of framing for still pictures which are to be projected is to mask out undesirable areas with masking tape or any suitable form of mask. Both inside edges are masked first at the edge of the picture on the film. In either view the distance from the image of the near point to the inside edge is measured and used to establish the distance from the image of the near point to the outside edge in the other view. To place the frame slightly in front of the near point this distance should be slightly reduced. This establishes the width of one picture. The other picture must be made the same width to insure a frontal frame. These steps are shown in Fig. 7.

The above is a procedure for framing individual slides and is not feasible for movies or for automatic framing of a large number of pictures. Such framing may be effectually done either in the camera so that the framing is on the film or at the screen.

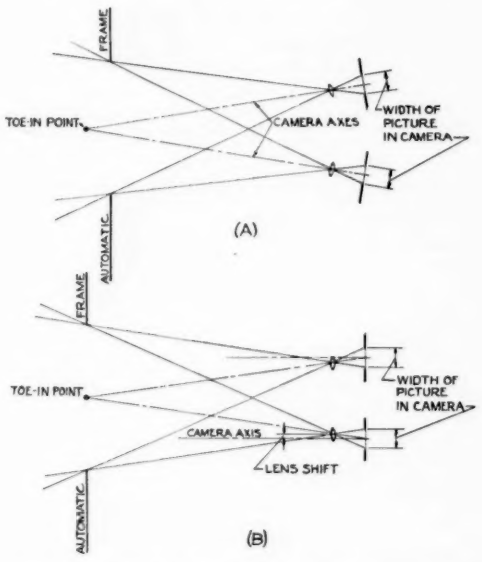


FIG. 8. (A) Framing on film by shifting camera lenses; (B) framing on film by use of toe-in.

Framing on the film may be achieved by using the natural edges of the picture as a frame or by building an adjustable frame directly in front of the focal plane of the camera. In the latter case two equal size openings smaller than the normal size of the picture and variable in separation would have to be built into the camera and their separation calibrated in terms of the distance to the point where the frame will appear. This distance would be the distance to the point whose images appeared exactly in the center of each opening. Such a framing device would require a specially constructed camera.

Two methods of framing by use of the natural edges of the picture are possible. Figure 8 illustrates these.

If the two axes of the camera are toed-in on any point the resultant frame will appear to lie very near the frontal plane containing this point.* This can be seen from Fig. 8(a). This is inherently an extremely simple method of framing. With a

* This method used for Chrysler movie at New York World's Fair. Taken by Loucks and Norling Studios, New York.

camera capable of toed-in axes, the frame for any scene can be set in advance permanently on the film. The setting can be made visually on a ground glass containing hair lines for the centers of the openings. The two hair lines need then only be set on the nearest object in the scene to place the frame there. This system has two advantages; first, it fixes the frame permanently so that untrained operatives cannot destroy it. Second, it utilizes the full width of the picture so that no edge area is lost. Exactly the same effective framing can be achieved in a double-film camera by building the camera so that the distance between lenses may be reduced without moving the film. This is shown in Fig. 8(b). This has the advantage over the toe-in method of avoiding any keystoning effect. It requires, however, a specially built camera.

Both of these methods demand that the projectors also be toed-in to achieve the proper separation between views at the screen. Pictures taken with a toe-in and projected parallel are, in effect, increased in separation and lead to divergence unless the scene is limited in depth as shown in formula (2).

Framing on the film has the very distinct advantage that the film may then be shown at any size or at any separation and the frame will remain in the same relative position with respect to the image.

If a film is already made without adequate frames, framing may theoretically be done on the film by blacking out the edges where necessary. This is a difficult problem on movie films. Furthermore it leads to different widths of picture for each different frame location.

Framing at the screen may be achieved in any projection in which the infinity points are not crossed, that is, in which the left infinity point is to the left of the right infinity point. This is always true in parallel projection. The framing is achieved by masking the sides of the screen with some black material to points inside the edges of both pictures.‡ This automatically places a frame at the plane of the screen for all pictures. It demands that all pictures be taken so that no object will appear forward of the plane of the screen, in other words the photographing must be done with this method of framing in mind. It further demands that the projection system be rigidly fixed so that the pictures will always be of the same size and in the same place on the screen.

‡ This method used for Pennsylvania Railroad movie at San Francisco World's Fair. Taken by Loucks and Norling Studios, New York.

Conclusions

Certain general observations on stereo projection can be made from all of the above discussion. The most important of these is the fact that the size of the screen should be known in advance when taking stereo pictures. This is a distinct drawback. It may be noted that, so long as projection is kept to small sizes, say 6 feet of width or less, this problem is not acute unless very small or near objects are photographed. The problem becomes more acute as the size of screen is increased. It points definitely toward a standardization on a few screen sizes in the future, as some leeway is permissible. In general, if pictures are taken for the largest screen that may be used they may be shown on smaller screens with some loss of depth.

It can also be seen that many problems, particularly of an interpretive nature, remain to be solved. Distortions have not been discussed. No careful statistical study has been made of the degree of the various types of distortion which are permissible without their becoming objectionable to the average audience. Far from enough is known of the eyestrain involved in the accommodation-convergence breakdown, in vertical errors, and in divergence.

It should be remembered, however, that the above problems are in the nature of refinements. The present stereoscopic movie, when intelligently taken and projected is a very good product, acceptable to even a critical observer. No form of still photography in any way approaches the reality, nor yields the satisfaction, of properly photographed and projected natural color stereographs.

Color Vision and the Autonomic Nervous System

S. V. KRAVKOV
The Helmholtz Institute of Ophthalmology, Moscow, U. S. S. R.
(Received January 3, 1941)

Quite different indirect stimuli often produce similar changes in the color sensitivity of the eye. It was assumed that this similarity of effect is due to all these indirect stimuli being similar as to their influence upon the autonomic nervous system. Direct experiments with instillation of adrenalin and pilocarpin into the eye have shown that color sensitivity does change when affected by these sympathico- and parasympathico-mimetic substances. In the light of the dependence of color vision on the autonomic nervous system, it is also possible to account for the seasonal fluctuations of color sensitivity recently described in special papers.

A NUMBER of experiments carried out during the last years in the physiological optics laboratory of The Helmholtz Institute of Ophthalmology in Moscow evidence the fact that the color sensitivity of the eye may change considerably in accordance with different accessory indirect stimuli acting upon the subject.[1]

These indirect stimuli applied in our laboratory were: sounds (780 and 2100 Herz, loudness from 25 to 95 decibels); smells (bergamot oil, geraniol); thermal stimulations (warming the hand with heated air); tastes (sugar).

The color sensitivity was established in the course of the dark adaptation both by a direct measurement of the threshold of the color sensitivity of the eye and by the method of critical flicker frequency.

The findings obtained by the latter method enable us to judge the color sensitivity of the eye, as in certain conditions the change of critical flicker frequency proceeds, as we have shown, in a direction inverse to that of the changes in color sensitivity.[2]

The experiments have shown that the indirect stimuli applied, different though they appeared to be, produced analogous changes in color sensitivity, these changes being as follows: Sensitivity to the rays of the green-blue part of the spectrum increases, that to red-orange rays diminishes, and that to the rays of the middle region of the spectrum (at about 570 mμ) and to its extreme violet and red rays undergoes no perceptible change.

This is to be seen from Fig. 1, showing the changes in the color sensitivity of the eye under the influence of sound stimulation, and the changes in critical flicker frequency under the influence of odor, thermal and taste stimuli. In the conditions of our experiments, the changes in critical flicker frequency, as has been said above, are contrary to the changes in color sensitivity, that is, an increase in critical flicker frequency implies a decrease of sensitivity and vice versa.

On the abscissa are marked the light waves in millimicrons; on the ordinate, the relative values of sensitivity or critical flicker frequency found during the action of the accessory stimulation. Value "1" on the ordinate represents the level of sensitivity (or of the critical flicker frequency) observed under the same conditions but in the absence of accessory stimulation. The data shown in Fig. 1 are the results of numerous experiments performed on several observers (2 to 5).

The similarity of the effect on color vision of acoustic, odor, thermal and taste accessory

[1] S. V. Kravkov, "The influence of sound upon the light and color sensitivity of the eye," Acta Ophthalmol. **14**, 348 (1936).
S. V. Kravkov, "The influence of acoustic stimulation upon the color sensitivity of a protanopic eye," Acta Ophthalmol. **15**, 337 (1937).
P. A. Yakovlev, "The influence of acoustic stimuli upon the limits of visual fields for different colors," J. Opt. Soc. Am. **28**, 286 (1938).
S. V. Kravkov, "The influence of the loudness of the indirect sound stimulus on the color sensitivity of the eye," Acta Ophthalmol. **17**, 324 (1939).
S. V. Kravkov, "The influence of odors upon color vision," Acta Ophthalmol. **17**, 426 (1939).
S. V. Kravkov, "Some new findings on color vision," Acta Medica URSS **2**, 461 (1939).
O. A. Dobriakova, "The influence of thermal and taste stimuli on color vision" (1939, unpublished).

[2] S. V. Kravkov, "Critical frequency of flicker and indirect stimuli," Comptes rendus (Doklady) de l'Académie des Sciences de l'URSS **22**, 64 (1939).

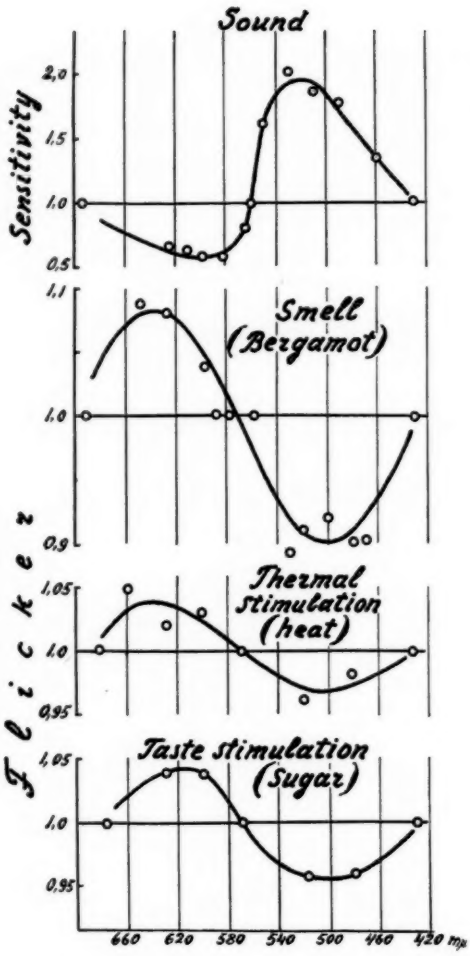


FIG. 1. Changes in color vision when affected by different indirect stimuli.

stimuli applied induced us to seek for some common features in the physiological nature of the influence of these stimuli upon the organism. Measurements carried out in our laboratory by Dr. Zaretzkaya have shown that all of the above-mentioned accessory stimuli quicken the pulse, that is, act sympathico-mimetically. The question naturally arose of a possible influence of our accessory stimuli upon the color sensitivity exerted by means of the changes they produce in the condition of the autonomic nervous system.

Modern physiologists, as is known, consider it highly probable that the autonomic nervous system exerts an adaptive-trophic influence upon our sense organs.[3] Changes in the sensitivity of scotopic (rod) vision when influenced by changes in the condition of the autonomic nervous system have already been described in literature.[4] We have, therefore, made some experiments to find out how color (cone) vision is influenced by substances affecting the organism in a way similar to the excitation of either of the two parts of the autonomic nervous system, that is, substances producing a sympathico-mimetic or parasympathico-mimetic effect. The substances of this kind applied by us were adrenalin (1 to 3 drops, solution 1 : 1000) and pilocarpin (2 to 3 drops, solution 1 : 100) instilled into the conjunctival sac of the observer's eye.[5] The experiments with adrenalin carried out on seven persons have shown that 30 to 40 minutes after its instillation into the eye, changes set in in the color sensitivity of the foveal vision, in all respects similar to those resulting from the action of the accessory stimuli described above. See the data of observer *S* shown in Fig. 2. On the ordinate are marked the relative values of color sensitivity as observed forty minutes after the instillation of adrenalin. The ordinate "1" stands for the sensitivity value observed under the same conditions of adaptation without adrenalin; on the abscissa are shown the light waves, sensitivity to which was to be determined, in millimicrons. The experiments with pilocarpin

[3] E. Brücke, "Fortschritte in der Erkenntnis des vegetativen Nervensystems" 90-te Versamml. d. Gesells. deutsch. Naturforscher u. Aertzte zu Hamburg (1928), p. 923.
L. A. Orbeli, "Lectures on the physiology of the nervous system," Leningrad, 1935, p. 319 (in Russian).
I. Tinel, *Le systeme nerveux végétatif* (Paris, 1937), Chapt. X.

[4] H. Rothhan, "Ueber die Beeinflussung der Netzhautfunktion durch Adrenalin" Klin. Monats. f. Augenheil. **75**, 747 (1925).
W. Hess and F. Lehmann, "Der Einfluss vegetativer Reizstoffe auf den Adaptierungszustand der Retina" Pflügers Arch. f. Physiol. **211**, 603 (1926).
H. Altenburger and F. Kroll, "Die vegetative Beeinflussung des optischen Systems" Zeits. f. Neurol. Psychiat. **124**, 527 (1930).
A. V. Lebedinsky, L. T. Zagorulko and I. P. Turzaev, "The influence of pain-stimulus upon the sensitivity of the dark-adapted eye," Physiol. J. USSR **16**, 740 (1933) (in Russian).

[5] S. V. Kravkov, "Ueber die Abhängigkeit unseres Farbensehens von dem vegetativen Nervensystem," "Beiträge zur physiologischen Optik" Gesammelte Abhand. herausg. Akad. Wiss. URSS, Vol. I (1941) (in Russian, with German summary).

performed on seven persons have shown that under its influence (40 to 50 minutes after the instillation) changes set in in the color vision which were contrary to those observed after the instillation of adrenalin, that is, sensitivity to orange-red rays increased, while sensitivity to green and blue ones diminished. Thus we may consider it an established fact that the color sensitivity of our central (cone) vision is dependent in a definite way on the condition of the autonomic nervous system.

A comparison of the curve of change in sensitivity under the influence of adrenalin (see Fig. 2) with the curves of the three fundamental color excitations of the eye as they were recently established in accordance with Helmholtz' theory of color vision,[6] brings us to the following conclusion: Changes in the autonomic nervous system directly affect only the green-sensitive apparatus of our vision. Consequently neither adrenalin nor pilocarpin produce any changes in the sensitivity to the extreme red (690 to 700 mμ)

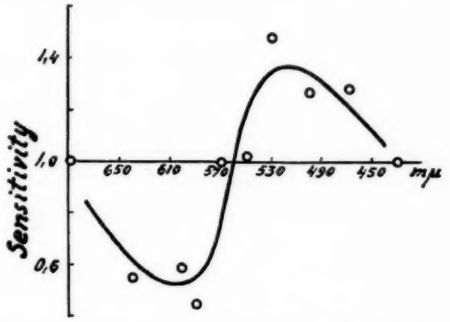


FIG. 2. Changes in color sensitivity following on the instillation of adrenalin.

or violet (430 mμ) rays of the spectrum that do not act upon the green-sensitive apparatus. The excitability of the green-sensitive apparatus of our vision increases when affected by a sympathico-mimetic substance (adrenalin), and decreases when affected by a parasympathico-mimetic one. When both the green-sensitive and the red-sensitive apparatuses of the eye are excited simultaneously, hyperexcitability of the green-sensitive apparatus is accompanied by hypoexcitability of the red-sensitive one.

A certain ratio between the excitations of the green-sensitive and the red-sensitive apparatuses admits of a reciprocal neutralization of these inverse changes as a result of which the sensitivity may remain unchanged. Such appears to be the case when the eye is affected by the yellow rays of the spectrum (at about 570 mμ).

The above experiments showing the dependence of color sensitivity on the autonomic nervous system throw some light upon the seasonal fluctuations of color sensitivity recently observed by Preston and also described by Dresler.[7] According to their findings, sensitivity to green rays is relatively higher in summer. It is to be assumed that this fact is due to a higher tonus of the sympathetic nervous system which, again, may be connected with seasonal changes in the functioning of the endocrine glands.

Thus the dependence of color sensitivity on the condition of the autonomic nervous system may account for many diverging data found in literature[8] dealing with the luminosity curve of our cone vision.

[6] N. T. Fedorov and V. F. Fedorova, "Study of color vision," Bull. Acad. Sci. URSS. Cl. Sci. math. et natur., p. 1431 (1935).

[7] A. Dresler, "Ueber eine jahreszeitliche Schwankung der spektralen Hellempfindlichkeit," Das Licht **10**, No. 4, p. 79 (1940).

[8] Kasson S. Gibson, "Spectral luminosity factors," J. Opt. Soc. Am. **30**, 51 (1940).

Unit Planes in an Optical System Including a Plane-Parallel Plate

K. PESTRECOV
Scientific Bureau, Bausch and Lomb Optical Company, Rochester, New York
(Received January 24, 1941)

IN solving some problems of optical design, it is sometimes more convenient to locate the unit planes of an optical system by the explicit algebraic formulas instead of by a paraxial ray tracing. These formulas for a single lens as well as for compound optical systems are generally known, and may be found in most of the text- and handbooks. It seems, however, that no such ready-to-use formulas are available in the case of a combination with a plane-parallel plate,—at least they could not be found in a number of books reviewed for this purpose by the author. The formulas in question could be easily derived by a paraxial tracing and a simple geometrical construction; the author was, however, induced by some practical considerations to approach the problem in a more general way. The following solution, although not the simplest one, seems to be interesting from the mathematical point of view, and may have an educational value as well.

The basic formulas used in this derivation are those of Hardy and Perrin[1] slightly modified and adapted to the sign convention of this article (distances are positive at the right and negative at the left of a reference point). It is assumed that the optical systems are surrounded by air and that the light is coming from the left.

UNIT PLANES OF A PLANE-PARALLEL PLATE

The lateral magnification is unity for any pair of conjugate points in the case of a plane-parallel plate. Therefore, any such pair may be considered as a pair of principal points and the plate in a compound system may be treated as a special case of a telescopic system in a compound system. The unit planes of the compound system may be then located by customary methods.[2] It is possible, however, to derive a definite location for the unit planes in a plane-parallel plate if the plate is considered as the limit of a simple lens, both radii of which increase to infinity. To simplify the derivation, the assumption is made that either $R_1=R_2$ or $R_1=-R_2$, and both cases are analyzed separately.

Case (a)—$R_1=R_2$

For any simple lens limited by spherical surfaces of equal curvature $(1/R)$ the following expressions hold for the focal length (f), for the distances of the unit planes from the corresponding vertices (V_1H and V_2H'), and for the separation of the principal points (HH'):

$$f=\frac{R^2n}{(n-1)^2t}; \qquad V_1H=-\frac{R}{n-1}; \qquad V_2H'=-\frac{R}{n-1} \quad \text{and} \quad HH'=t. \tag{1}$$

As R approaches $\pm\infty$, these formulas yield:

$$f=+\infty; \quad V_1H=\mp\infty; \quad V_2H'=\mp\infty; \quad HH'=t.$$

Thus it is seen that if a plane-parallel plate is considered as the limit of a lens with $R_1=R_2\rightarrow\infty$, its unit planes are at infinity, being separated by t (the thickness of the plate).

If this conception of a plane-parallel as a limit is correct, the formula $1/s'-1/s=1/f$, where s' and s are distances referred to the unit planes, should yield the same position of the image as is usually derived for a plane-parallel plate; also the magnification m must be unity independent from the position of the object. The position of the image formed by a plane-parallel plate is defined by the following formulas:

$$v=u-t/n \quad \text{or} \quad SS'=t-t/n,$$

where v and u are the image and object distances referred, respectively, to the second and first plate surface, and SS' is the linear shift of the image from the object (in the direction of the incident light).

Now it may be written from Fig. 1 and Eq. (1):

$$s=HV_1+u=\frac{R}{n-1}+u \quad \text{and} \quad s'=\frac{R}{n-1}+v.$$

[1] Arthur C. Hardy and Fred H. Perrin, *The Principles of Optics* (McGraw-Hill Company, 1932), pp. 58–61.
[2] H. T. Flint, *Geometrical Optics* (The Sherwood Press, 1937), pp. 110, 116–128.

Then

$$m=\frac{s'}{s}=\frac{R/(n-1)+v}{R/(n-1)+u}; \quad \text{and} \quad \lim_{R\to\pm\infty} m=1$$

for any finite u and v. Also

$$\frac{1}{R/(n-1)+v}-\frac{1}{R/(n-1)+u}=\frac{(n-1)^2t}{R^2n}.$$

After obvious transformations, we obtain:

$$(n-1)v=\frac{R^2n(n-1)u-R^2(n-1)t-R(n-1)^2tu}{[R+(n-1)u](n-1)t+R^2n};$$

reducing by R^2, we find, as R increases to infinity:

$$\lim v=u-t/n \quad \text{or} \quad \lim SS'=t-t/n.$$

These expressions are identical with the paraxial formulas.

Case (b)—$R_1=-R_2(=R)$

For any simple lens whose radii satisfy this condition, the following expressions hold:

$$f=\frac{R^2n}{(n-1)(2R+(n-1)t/n)},$$

$$V_1H=\frac{Rt}{2nR-(n-1)t},$$

$$V_2H'=-\frac{Rt}{2nR-(n-1)t},$$

$$HH'=t-\frac{t}{n-(n-1)t/2R}.$$

As R approaches $\pm$ infinity, these expressions become:

$$f=\pm\infty; \quad V_1H=t/2n \quad V_2H'=-t/2n \quad \text{and} \quad HH'=t-t/n. \tag{2}$$

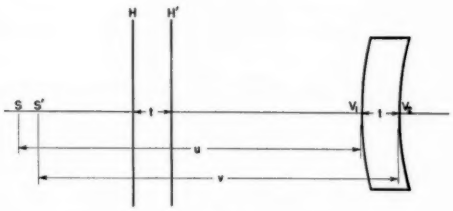


FIG. 1.

In this case the unit planes are located inside the plane-parallel plate at equal distances from its surfaces. It should be noted here that similar formulas were derived by Hastings;[3] however, his derivation is not sufficiently rigorous.

In this case, as long as the object is at a finite distance from the plate, s and s' are also finite; therefore, it may be immediately written:

$$\lim_{f\to\pm\infty}(1/s'-1/s)=0; \quad \text{or} \quad s'=s \quad \text{and} \quad m=1.$$

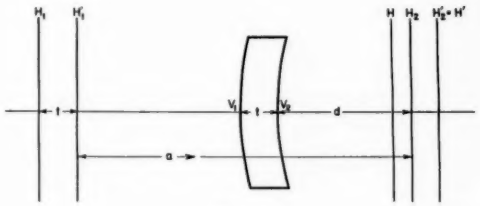


FIG. 2.

Then

$$SS'=-s+(t-t/n)+s'=t-t/n \quad \text{and} \quad v=u-t/n.$$

Thus, both solutions (a) and (b), although resulting in apparently different positions for the unit planes, are equivalent, yielding the same position of the image and magnification (unity) as the paraxial formula for the plane-parallel plate.

It will be now shown that a combination of either case with another optical system results in identical compound systems.

COMBINATION OF TWO OPTICAL SYSTEMS OF WHICH ONE IS A PLANE-PARALLEL PLATE

Case (a)—$R_1=R_2$

In Fig. 2 the unit planes H_1 and H_1' are those of the lens V_1V_2 which becomes a plane-parallel plate as $R_1=R_2$ increases to infinity. H_2 and H_2' are the unit planes of another optical system of a finite f_2. H and H' are the unit planes of the compound system. It may be written from Fig. 2:

$$a=H_1'V_2+d.$$

Now the distance of the second unit plane of the second system from the second unit plane of

[3] Charles S. Hastings, *New Methods in Geometrical Optics* (Macmillan, 1927), p. 33.

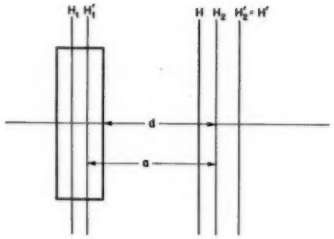


FIG. 3.

the combination is given by the formula:

$$II'H_2'=\frac{af_2}{f_1+f_2-a}.$$

Substituting in this formula the value a as given above and $H_1'V_2$ and f_1 from Eq. (1), we obtain:

$$II'H_2'=\frac{(R/(n-1)+d)f_2}{R^2n/(n-1)^2t+f_2-R/(n-1)-d};$$

dividing the numerator and denominator by R, we find that

$$\lim_{R\to\pm\infty} II'H_2'=0.$$

The second unit plane of the combination coincides with the second unit plane of the second system.

From the same figure, we have:

$$HH_2=a-H_1'H=a-(H_1H-t)$$
$$=H_1'V_2+d-H_1H+t.$$

By substituting

$$H_1H=\frac{af_1}{f_1+f_2-a}$$

and the values of $H_1'V_2$ and f_1 as before, we obtain after obvious transformations:

$$HH_2=d+t+\frac{R[(f_2-d)(n-1)t-Rt]-ndR^2}{nR^2+(f_2-d)(n-1)^2t-R(n-1)t}.$$

Reducing the numerator and the denominator by R^2, we obtain:

$$\lim_{R\to\pm\infty} HH_2=d+t-(t+nd)/n=t-t/n.$$

The first unit plane of the combination is shifted from the first unit plane of the second system by $t-t/n$ toward the plate.

Case (b)—$R_1=-R_2$

In this case the optical separation a is finite as long as the plate is at a finite distance from the second system and, therefore, it may be immediately written from Fig. 3 and Eq. (2).

$$a=d+t/2n.$$

Then

$$\lim_{f_1\to\pm\infty} II'H_2'=\lim\frac{af_2}{f_1+f_2-a}=0.$$

And

$$HH_2=a+H_1H_1'-H_1H=a+t-\frac{t}{n}-\frac{af_1}{f_1+f_2-a}.$$

As f_1 approaches $\pm$ infinity,

$$\lim HH_2=a+t-t/n-a=t-t/n.$$

The solution in the case (b) is identical with that in the case (a). It will be noted that the positions of the unit planes of the compound system are independent of the position of the plane-parallel plate (as long as it remains in the same optical space)—the position term d does not appear in the final formulas—which is equivalent to the previous statement, that, generally, any two conjugate points in the case of a plane-parallel plate may be considered as a pair of its principal points (unit planes).

As in any other compound system, the positions of the unit planes derived for the system under consideration are correct only as long as the compound system remains a physical unit. An interesting illustration of this fact may be made on the lens bench. A lens is mounted on the nodal slide of the bench and its second nodal point (unit plane) is found, as usual—by locating the axis about which rotation does not cause a displacement of the image observed with a microscope. If a plane-parallel plate is now attached to the lens on its side facing the microscope, the image of the target formed by the compound system will no longer be in the focus of the microscope, and it will also be found, by rotation of the entire combination, that the vertical axis of the slide no more coincides with the second nodal point (unit plane) of the new optical system. This indicates that the second unit plane of the compound system is shifted away from the second unit plane previously located for the lens itself. By moving the lens

(with the attached plate) relative to the vertical axis of the slide by $t-t/n$ away from the microscope, the image will be brought again into focus and the second unit plane of the combination will be brought into coincidence with the vertical axis of the slide. Since the lower part of the nodal slide remains stationary, the focal length read on the bench scale will be the same as for the lens alone. If, however, the plate is placed separate from the lens (between the lens and the microscope), it will be observed, by rotating the lens, that the second unit plane of the lens still coincides with the axis of the slide, although the image will not be in focus. In this case, the plane-parallel plate should be considered as a part of the optical system of the microscope, in consequence of which the focal point (and the first unit plane) of this compound system is shifted toward the lens by $t-t/n$. To bring the image again in focus, without changing the reading for the focal length of the lens, it will be necessary to recalibrate the lens bench by shifting the microscope away from the lens. It may be noted that placing the plane-parallel plate in front of the lens (between the lens and collimator) would have no effect on the lens bench observations because then the first unit plane would be shifted, but the second unit plane of the combination would coincide with that of the lens.

Summary

Considering the plane-parallel plate as the limit of a simple lens with both its radii increasing to infinity, a definite location may be assigned to the unit planes of the plate.

If the absolute values of the radii are assumed to be equal, two solutions are possible, depending upon whether the radii are of the same or different signs.

In the first case, the unit planes are found to be at infinity, separated by t—the thickness of the plate.

In the second case, the unit planes are located inside the plate separated by $t-t/n$.

Both solutions yield unity for lateral magnification, independent of positions of the object and yield the same position of the image as the paraxial formula. If plane-parallel plate is combined with another optical system, both solutions yield identical positions for the unit planes of the compound system. It is found that in the resultant system the unit plane of the space in which the plate is located is shifted from the corresponding unit plane of the second component by $t-t/n$ in the direction toward the plate; the other unit plane of the compound system coincides with the corresponding plane of the second component.

Optical Society of America, Incorporated

Notice of Election of Officers

In accordance with Article III, Section 6 of the By-Laws of the Optical Society of America the following officers and directors are to be elected at or before the Annual Meeting in October, 1941:

President, Vice President — for a term of two years

Secretary, Treasurer, 3 Members-at-large of the Board of Directors — for a term of four years

The present officers are:

	Term expires at annual meeting in
President—K. S. Gibson	1941
Vice President—A. G. Worthing	1941
Secretary—A. C. Hardy	1941
Treasurer—H. F. Kurtz	1941
Past President—R. C. Gibbs	1941
Editor—G. R. Harrison	1944
(Elected by Directors)	
Directors-at-large	
H. P. Gage	1941
W. F. Meggers	1941
Brian O'Brien	1941
J. W. Beams	1943
Rudolf Kingslake	1943

In accordance with the By-Laws, Article III, Section 4:

> "No member at large of the board of directors shall be eligible for re-election."

H. P. Gage, W. F. Meggers, and Brian O'Brien are not eligible for re-election.

The following Committee on Nominations has been appointed and confirmed in accordance with Article III, Section 6 of the By-Laws:

R. C. Gibbs, *Chairman*
Cornell University,
Ithaca, New York.

W. E. Forsythe,
Nela Park,
Cleveland, Ohio.

L. V. Foster,
58 Somershire Drive,
Rochester, New York.

C. E. Foss,
1 Washington Mews,
New York, New York.

H. T. Wensel,
Nat. Bureau of Standards,
Washington, D. C.

It will be the duty of this committee to make nominations for the above-mentioned election. The Committee report is to be made not later than July 15, 1941.

Although, after the committee has made its report, an opportunity will be given for additional nominations by petition, members are urged, as soon as possible, to suggest names which should be considered by the committee. Suggestions may be sent to the Chairman or any other member of the Committee.

Only regular and honorary members are eligible to election. A list of members, officers and past officers was published in the J. Opt. Soc. Am., Vol. 28, No. 6, pp. 190–208, June, 1938. (Later transfers to regular membership are given in the Secretary's "Annual Report" published in the J. Opt. Soc. Am. Vol. 29, No. 6, pp. 248–249, June, 1939; the Secretary's "Annual Report" published in the J. Opt. Soc. Am., Vol. 30, No. 6, pp. 259–260, June, 1940; and in the "Minutes of the Sixteenth Annual Meeting of the Directors of the Optical Society" held on October 22, 1940, and published on page 650 of the J. Opt. Soc. Am., Vol. 30, No. 12, December, 1940.)

Arthur C. Hardy, *Secretary*